BUNDESREPUBLIK DEUTSCHLAND

Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen: 199 32 914.1

Anmeldetag: 14. Juli 1999

Anmelder/Inhaber: BASF Aktiengesellschaft, Ludwigshafen/DE

Bezeichnung: Corynebacterium glutamicum genes encoding proteins involved in genetic stability, gene expression, and protein secretion and folding

IPC: C 07 K 14/34

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 17. Juli 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN GENETIC STABILITY, GENE EXPRESSION, AND PROTEIN SECRETION AND FOLDING

Abstract of the Disclosure

Isolated nucleic acid molecules, designated SES nucleic acid molecules, which encode novel SES proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing SES nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated SES proteins, mutated SES proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of SES genes in this organism.

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN GENETIC STABILITY, GENE EXPRESSION, AND PROTEIN SECRETION AND FOLDING

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Summary of the Invention

This invention provides novel nucleic acid molecules which may be used to identify or classify *Corynebacterium glutamicum* or related species of bacteria. *C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The nucleic acid molecules therefore can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. While *C. glutamicum* itself is nonpathogenic, it is related to other Corynebacterium species, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria), which are important human pathogens. The ability to identify the presence of Corynebacterium species therefore also can have significant clinical relevance, e.g., diagnostic applications. Further, these nucleic acid molecules may serve as reference points for the mapping of the *C. glutamicum* genome, or of genomes of related organisms.

These novel nucleic acid molecules encode proteins, referred to herein as genetic stability, gene expression, or protein secretion/folding (SES) proteins. These SES proteins are capable of, for example, performing a function involved in the repair or recombination of DNA, transposition of genetic material, expression of genes (i.e.,

involved in transcription or translation), protein folding, or protein secretion in *Corynebacterium glutamicum*. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey et al., U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al. *J. Bacteriol.* 162: 591-597 (1985); Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984); and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an SES protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, modulation of proteins involved directly in transcription or translation (e.g., polymerases or ribosomes) such that they are increased in number or in activity should increase global cellular transcription or translation (or rates of these processes). This increased cellular gene expression should include those proteins involved in fine chemical biosynthesis, so an increase in yield, production, or efficiency of production of one or more desired compounds may occur. Modifications to the transcriptional/ translational protein machinery of *C. glutamicum* such that the regulation of these proteins is altered may also permit increased expression of genes involved in the production of fine chemicals. Modulation of the activity or number of proteins involved in polypeptide folding may permit an increase in the overall production of correctly folded molecules in the cell, thereby increasing the possibility that desired proteins (e.g., fine chemical biosynthetic proteins) are able to function properly. Further, by mutating proteins involved in secretion from *C. glutamicum* such that they are increased in number or activity, it may be possible to increase the secretion of a fine chemical (e.g., an enzyme) from cells in fermentor culture, where it may be readily recovered.

Genetic modification of the SES molecules of the invention may also result in indirect modulation of production of one or more fine chemicals. For example, by increasing the number or activity of a DNA repair or recombination protein of the invention, one may increase the ability of the cell to detect and repair DNA damage. This should effectively increase the ability of the cell to maintain a mutated gene within its genome, thereby increasing the likelihood that a transgene engineered *into C. glutamicum* (e.g., encoding a protein which will increase biosynthesis of a fine chemical) will not be lost during culture of the microorganism. Conversely, by

decreasing the number or activity of one or more DNA repair or recombination proteins, it may be possible to increase the genetic instability of the organism. Such manipulations should improve the ability of the organism to be modified by mutagenesis without the introduced mutation being corrected. The same holds true for proteins involved in transposition or rearrangement of genetic elements in *C. glutamicum* (e.g., transposons). By mutagenizing these proteins such that they are either increased or decreased in number or activity, it is possible to simultaneously increase or decrease the genetic stability of the microorganism. This has a profound impact on the ability of any other mutation to be introduced into *C. glutamicum*, and on the ability of introduced mutations to be retained. Transposons also offer a convenient mechanism by which mutagenesis of *C. glutamicum* may be performed; duplication of desired genes (e.g., fine chemical biosynthetic genes) is readily accomplished by transposon mutagenesis, as is disruption of undesired genes (e.g., genes encoding proteins involved in degradation of desired fine chemicals).

By modulating one or more proteins (e.g. sigma factors) involved in the regulation of transcription or translation in response to particular environmental conditions, it may be possible to prevent the cell from slowing or stopping protein synthesis under unfavorable environmental conditions, such as those found in large-scale fermentor culture. This should lead to increased gene expression, which in turn may permit increased biosynthesis of desired fine chemicals under such conditions. Mutagenesis of proteins involved in protein secretion systems may result in modulated secretion rates. Many such secreted proteins have functions critical for cell viability (e.g., cell surface proteases or receptors). An alteration of a secretory pathway such that these proteins are more readily transported to their extracellular location may improve the overall viability of the cell, and thus result in greater numbers of *C. glutamicum* cells capable of producing fine chemicals during large-scale culture. Further, the secretion apparatus (e.g., the sec system) is also known to be involved in the insertion of integral membrane proteins (e.g., pores, channels, or transporters) into the membrane. Thus, the modulation of activity of proteins involved in protein secretion from *C. glutamicum* may affect the ability of the cell to excrete waste products or to import necessary metabolites. If the activity of these secretory proteins is increased, then the ability of the cell to produce fine chemicals may be similarly increased. If the activity of these secretory proteins is decreased, then there may be insufficient nutrients available to support overproduction of desired compounds, or waste products may interfere with such biosynthesis.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as SES proteins, which are capable of, for example, participating in

the repair or recombination of DNA, transposition of genetic material, expression of genes (i.e., the processes of transcription or translation), protein folding, or protein secretion in *Corynebacterium glutamicum*. Nucleic acid molecules encoding an SES protein are referred to herein as SES nucleic acid molecules. In a preferred embodiment, an SES protein participates in improving or decreasing genetic stability in *C. glutamicum*, in the expression of genes (i.e., in transcription or translation) or protein folding in this organism, or in protein secretion from *C. glutamicum*. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs) comprising a nucleotide sequence encoding an SES protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of SES-encoding nucleic acid (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred SES proteins of the present invention also preferably possess at least one of the SES activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g., sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an SES activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which

is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (e.g., an SES fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C glutamicum* SES protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an SES protein by culturing the host cell in a suitable medium. The SES protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an SES gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated SES sequence as a transgene. In another embodiment, an endogenous SES gene within the genome of the microorganism has been altered, e.g., functionally disrupted, by homologous recombination with an altered SES gene. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

Still another aspect of the invention pertains to an isolated SES protein or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated SES protein or portion thereof can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of

transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. In another preferred embodiment, the isolated SES protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*.

The invention also provides an isolated preparation of an SES protein. In preferred embodiments, the SES protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated SES protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated SES protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98,%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of SES proteins also have one or more of the SES bioactivities described herein.

The SES polypeptide, or a biologically active portion thereof, can be operatively linked to a non-SES polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the SES protein alone. In other preferred embodiments, this fusion protein participates in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an SES nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an SES nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates SES protein activity or SES nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* processes involved in genetic stability, gene expression, protein folding, or protein secretion such that the yield, production, or efficiency of production of a desired fine chemical by this microorganism is improved. The agent which modulates SES protein activity can be an agent which stimulates SES protein activity or SES nucleic acid expression. Examples of agents which stimulate SES protein activity or SES nucleic acid expression include small molecules, active SES proteins, and nucleic acids encoding SES proteins that have been introduced into the cell. Examples of agents which inhibit SES activity or expression include small molecules and antisense SES nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant SES gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

Detailed Description of the Invention

The present invention provides SES nucleic acid and protein molecules which are involved in the repair or recombination of DNA, in the transposition of genetic

material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (e.g., where overexpression or optimization of activity of a protein involved in secretion of a fine chemical (e.g., an enzyme) has a direct impact on the yield, production, and/or efficiency of production of a fine chemical from the modified *C. glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (e.g., where modulation of the activity or number of copies of a *C. glutamicum* DNA repair protein results in alterations in the ability of the microorganism to maintain the introduced mutation, which in turn may impact the production of one or more fine chemicals from such a strain). Aspects of the invention are further explicated below.

I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g., hyaluronic acid and trehalose), aromatic compounds (e.g., aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosynthesis, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino aids – technical production and use, p. 466-502 in Rehm et al. (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E.(1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α-ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β-carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B_1) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B_2) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B_6' (e.g., pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate

biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5`-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B_5), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5`-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B_{12}) and porphyrines belong to a group of chemicals characterized by a tetrapyrole ring system. The biosynthesis of vitamin B_{12} is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B_6, pantothenate, and biotin. Only Vitamin B_{12} is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language

"purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (i.e., AMP) or as coenzymes (i.e., FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (e.g. Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (e.g., thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (e.g., ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (e.g., IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and Related Compounds in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from

ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in α, α-1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto et al., (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) J. Japan 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

II. Genetic Stability: Protein Synthesis and Protein Secretion in *C. glutamicum*

The production of a desired compound from a cell such as *C glutamicum* is the culmination of a large number of separate yet interrelated processes, each of which is critical to the overall production and release of the compound from the cell. In engineering a cell to overproduce one or more fine chemicals, consideration must be given to each of these processes to ensure that the biochemical machinery of the cell will be compatible with such genetic manipulation. Cellular mechanisms of particular importance include the stability of the altered gene(s) upon introduction into the cell, the ability of the mutated gene to be properly transcribed and translated (including issues of codon usage), and the ability of the mutant protein product to be appropriately folded and/or secreted.

A Bacterial Repair and Recombination Systems

Cells are constantly exposed to nucleic acid-damaging agents, such as UV irradiation, oxygen radicals, and alkylation. Further, even the action of DNA polymerases is not error-free. Cells must maintain a balance between genetic stability (which ensures that genes necessary for vital cellular functions are not damaged during normal growth and metabolism) and genetic variability (which permits cells to adapt to a changing environment). Therefore, there exist separate, but interrelated pathways of DNA repair and DNA recombination in most cells. The former serves to stringently correct errors in DNA molecules by either directly reversing the damage or excising the damaged region and replacing it with the correct sequence. The latter recombination system also repairs nucleic acid molecules, but only those lesions that result in damage to both strands of DNA such that neither strand is able to serve as a template to correct the other. Recombination repair and the SOS response may readily lead to inversions, deletions, or other genetic rearrangements within or around the region of the damage, which in turn promotes a certain degree of genomic instability which may contribute to the ability of the cell to adapt to changing environments or stresses.

High-fidelity repair mechanisms include direct reversal of DNA damage and excision of damage and resynthesis using the information encoded on the opposite DNA strand. Direct reversal of damage requires an enzyme having an activity opposite of that which originally damaged the DNA. For example, inappropriate methylation of DNA may be corrected by the action of DNA repair methyltransferases, and nucleotide dimers created by UV irradiation may be fixed by the activity of deoxyribodipyrimidine photo-lyase, which, in the presence of light, cleaves the dimer back to its constituent nucleotides (see Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, Wiley: New York, and references therein).

Precise repair of more extensive damage requires specialized repair mechanisms. These include the mismatch repair and excision repair systems. Damage to a single base may be corrected by a series of cleavage reactions, where first the sugar-base bond is cut, followed by cleavage of the DNA backbone at the site of damage and removal of the damaged base itself. Finally, DNA polymerase and DNA ligase act to fill in and seal the gap using the second DNA strand as a template. More significant DNA damage which results in altered conformation of the double helix is corrected by the ABC system, in which helicase II, DNA polymerase I, UvrA, UvrB, and UvrC proteins combine to nick the double helix at the site of damage, to unwind the damaged region in an ATP-dependent fashion, to excise the damaged region, and to fill in the missing region using the other strand as a template. Lastly, DNA ligase seals the nick. Specific repair systems also exist for G·T mismatches (involving the Vsr protein) and for small

deletion/insertion errors resulting in mispairing of the two strands (involving the methylation-directed pathway).

There also exist low-fidelity repair systems which are generally used to correct very extensive DNA damage in bacteria. Double-strand repair and recombination occurs in the presence of a lesion which affects both strands of DNA. In this situation, it is impossible to repair the damage utilizing the other strand as the template. Thus, this repair system involves a double-crossover event between the area of the lesion and another copy of the region on a homologous DNA molecule. This is possible because bacteria divide so rapidly that a second copy of genomic DNA is usually available before actual cell division occurs. This crossover event may readily lead to inversions, duplications, deletions, insertions and other genetic rearrangements, and thus increases the overall genetic instability of the organism.

The SOS response is activated when sufficient damage is present in the DNA that DNA polymerase III stalls and cannot continue replication. Under these circumstances, single-stranded DNA is present. The RecA protein is activated by binding to single-stranded DNA, and this activated form results in the activation of the LexA repressor, thereby lifting the transcriptional block of more than 20 genes, including UvrA, UvrB, UvrC, helicase II, DNA pol III, UmuC, and UmuD. The combined activities of these enzymes results in sufficient filling of the gap region that DNA pol III is able to resume replication. However, these gaps have been filled in with bases which should not be present; thus, this type of repair results in error-prone repair, contributing to overall genetic instability in the cell.

B. Transposons

The aforementioned systems, whether high or low fidelity, exist to repair DNA damage. In certain circumstances, this repair may accidentally incorporate additional genetic rearrangements. Many bacterial cells also have mechanisms specifically designed to cause such genetic rearrangements. Particularly well-known examples of such mechanisms are the transposons.

Transposons are genetic elements which are able to move from one site to another either within a chromosome or between a piece of extrachromosomal DNA (e.g., a plasmid) and a chromosome. Transposition may occur in multiple ways; for example, the transposable element may be cut out from the donor site and integrated into the target site (nonreplicative transposition), or the transposable element may alternately be duplicated from the donor site to the target site, yielding two copies of the element (replicative transposition). There is generally no sequence relationship between the donor and target sites.

There are a variety of results possible from such a transposition event. The integration of a transposable element into a gene disrupts the gene, usually abrogating its function entirely. An integration event that occurs in the DNA surrounding a gene may not perturb the coding sequence itself, but can have a profound effect on the regulation of the gene and thus, on its expression. Recombination events between two copies of a transposable element found in different portions of the genome may result in deletions, duplications, inversions, transpositions, or amplifications of segments of the genome. It is also possible for different replicons to fuse.

The simplest transposon-like genetic elements are termed insertion (IS) elements. IS elements contain a nucleotide region of varying length (though usually less than 1500 bases) lacking any coding regions, surrounded by inverted repeats at either end. Thus, since the IS element does not encode any proteins whose activity may be detected, the presence of an IS element is generally only observed due to a loss of function of one or more genes in which the IS element is inserted.

Transposons are mobile genetic elements which, unlike IS elements, contain nucleic acid sequences bounded by repeats which may encode one or more proteins. It is not unusual for these repeat regions to consist of IS elements. The proteins encoded by the transposon are typically transposases (proteins which catalyze the movement of the transposon from one site to another) and antibiotic resistance genes. The mechanisms and regulation of transposable elements are well known in the art and are have been described at least in, for example, Lengeler et al. (1999) Biology of Prokaryotes, Thieme Verlag: Stuttgart, p. 375-361; Neidhardt et al. (1996) *Escherichia coli* and *Salmonella*, ASM Press: Washington, D.C.; Sonenshein, A.L. et al., eds. (1993), *Bacillus subtilis*, ASM Press: Washington, D.C.; Voet, D. and Voet, J.G. (1992) Biochemie, VCH: Weinheim, p. 985-990; Brock, T.D., and Madigan, M.T. (1991) Biology of Microorganisms, 6th ed., Prentice Hall: New York, p. 267-269; and Kleckner, N. (1990) "Regulation of transposition in bacteria", *Annu. Rev. Biochem.* 61: 297-327.

C. Transcription

Gene expression in bacteria is regulated mainly at the level of transcription. The transcriptional apparatus consists of a number of proteins that can be divided into two groups: RNA polymerase (the processive DNA-transcribing enzyme) and sigma factors (which regulate gene transcription by directing RNA polymerase to specific promoter-DNA sequences which these factors recognize). The combination of RNA polymerase and sigma factors creates the RNA polymerase holoenzyme, an activated complex. Gram positive bacteria such as Corynebacteria contain only one type of RNA-polymerase, but a variety of different sigma factors specific for different promoters.

growth phases, environmental conditions, substrates, oxygen levels, transport processes, and the like, which permits adaptability of the organism to different environmental and metabolic conditions.

Promoters are specific DNA sequences that serve as docking sites for the RNA polymerase holoenzyme. Many promoter elements possess conserved sequence elements that may be recognized through homology searches; alternately, promoter regions for a particular gene may be identified using standard techniques such as primer extension. Many promoter regions from gram-positive bacteria are known (see, e.g., Sonenshein, A.L., Hoch, J.A., and Losick, R., eds. (1993) *Bacillus subtilis*, ASM Press: Washington, D.C.).

Promoter transcriptional control is influenced by several mechanisms of repression or activation. Specific regulatory proteins which bind promoters have the ability to block (repressors) or to assist (activators) the binding of the RNA holoenzyme, and thus to regulate transcription. The binding of these repressor and activator molecules in turn is regulated by their interactions with other molecules, such as proteins or other metabolic compounds. Transcription may alternately be regulated by factors influencing processes such as elongation or termination (see, e.g., Sonenshein, A.L., Hoch, J.A., and Losick, R., eds. (1993) *Bacillus subtilis*, ASM Press: Washington, D.C.). The ability to regulate transcription of genes in response to a variety of environmental or metabolic cues affords cells the ability to tightly control when a gene may be expressed and or how much of a gene product may be present in the cell at one time. This in turn prevents unnecessary expenditure of energy or unnecessary utilization of possibly scarce intermediate compounds or cofactors.

D. Translation and tRNA-Aminoacyl Synthetases

Translation is the process by which a polypeptide is synthesized from amino acids according to the information contained within an mRNA molecule. The main components of this process are ribosomes and specific initiation or elongation factors, such as IF1-3, EF-G, and EFTu (see, e.g., Sonenshein, A.L., Hoch, J.A., Losick, R., eds. (1993) *Bacillus subtilis*, ASM Press: Washington, D.C.).

Each codon of the mRNA molecule encodes a particular amino acid. The conversion from mRNA to amino acid is effected by transfer RNA (tRNA) molecules. These molecules consist of a single strand of RNA (between 60 and 100 bases), which exists in an L-shaped three dimensional structure having protruding areas, or 'arms'. One such arm forms base pairs with a particular codon sequence on the mRNA molecule. A second arm interacts specifically with a particular amino acid (the one encoded by the codon). Other arms of the tRNA include the variable arm, the TψC arm

(which bears thimidylate and pseudouridylate modifications), and the D arm (which bears a dihydrouridine modification). The function of these latter structures remains unknown, but their conservation between tRNA molecules suggests a role in protein synthesis.

In order for the nucleic acid-based tRNA molecule to associate with the correct amino acid, a family of enzymes, termed the aminoacyl-tRNA synthetases, must act. There exist many different of these enzymes, each of which is specific for a particular tRNA and a particular amino acid. These enzymes link the 3' hydroxyl of the terminal tRNA adenosine ribose moiety to the amino acid in a two step reaction. First, the enzyme is activated by reaction with ATP and the amino acid to result in an aminoacyl-tRNA synthetase-aminoacyl adenylate complex. Second, the aminoacyl group is transferred from the enzyme to the target tRNA where it remains in the high-energy state. Binding of the tRNA molecule to its cognate codon on the mRNA molecule then brings the high-energy amino acid attached to the tRNA into contact with the ribosome. Within the ribosome, the amino-acid charged tRNA (aminoacyl-tRNA) occupies one binding site (the A site) adjacent to a second site (the P site) containing a tRNA molecule whose amino acid arm is attached to the nascent polypeptide chain (peptidyl-tRNA). The activated amino acid on the aminoacyl-tRNA is sufficiently reactive that a peptide bond spontaneously forms between this amino acid and the next amino acid on the nascent polypeptide chain. Hydrolysis of GTP provides the energy for the transfer of the now-polypeptide chain-loaded tRNA from the A site to the P site of the ribosome, and the process repeats until a stop codon is reached.

There are a number of different steps at which translation may be regulated. These include the binding of the ribosome to mRNA, the presence of mRNA secondary structure, codon usage, or the abundance of particular tRNAs. Also, special regulation mechanisms such as attenuation may act at the level of translation. For an in-depth review of many of these mechanisms, see, e.g., Vellanoweth, R.L. (1993) "Translation and its Regulation" in: *Bacillus subtilis* and other Gram Positive Bacteria, Sonenshein, A.L. et al., eds., ASM Press: Washington D.C., p. 699-711, and references cited therein.

E. Protein Folding and Secretion

Synthesis of proteins by the ribosome results in polypeptide chains, which must take on a three-dimensional form before the protein can function normally. This three-dimensional structure is achieved by a process of folding. Polypeptide chains are flexible, and (in principle) move readily and freely in solution until they attain a conformation which results in a stable three-dimensional structure. However, it is sometimes difficult for proteins to fold correctly, either due to environmental conditions

(e.g., high temperature, where the extra kinetic energy present in the system makes it more difficult for the polypeptide to settle in the energy well of a stable structure) or due to the nature of the protein itself (e.g., the hydrophobic regions in nearby polypeptides have a tendency to aggregate and thereby sequester themselves from aqueous solution).

Proteinaceous factors have been identified that are able to catalyze, chaperone, or otherwise assist in the folding of proteins being synthesized either co- or posttranslationally. Members of these protein folding molecules are the prolyl-peptidyl isomerases (e.g., trigger factor, cyclophilin, and FKBP homologs), and proteins of the heat shock protein group (e.g., DnaK, DnaJ, GroEL, small heat shock proteins, HtpG and members of the Clp family (e.g., ClpA, ClpB, ClpW, ClpP, and ClpX)). Many of these proteins are essential for the viability of cells: in addition to their functions in protein folding, translocation, and processing, they frequently serve as key targets for the overall regulation of protein synthesis (see, e.g., Bukau, B., (1993) *Molecular Microbiology* 9(4): 671-680; Bukau, B., and Horwich, A.L. (1998) *Cell* 92(3):351-366; Hesterkamp, T., Bukau, C. (1996) *FEBS Lett.* 389(1):32-34; Yaron, A., Naider, F. (1993) *Critical Reviews in Biochemistry and Molecular Biology* 28(1):31-81; Scheibel, R., Buchner, J. (1998) *Biochemical Pharmacology* 56(6):675-682; Ellis, R.J., Hartl, F.U. (1996) *FASEB Journal* 10(1): 20-26; Wawrzynow, A. et al. (1996) *Molecular Microbiology* 21(5): 895-899; Ewalt, K.L., et al. (1997) *Cell* 90(3): 491-500).

Chaperones identified thus far function in one of two ways: they either bind and stabilize polypeptides, or they provide an environment in which folding may occur without interference. The former group, including, e.g., DnaK, DnaJ, and the heat shock proteins, bind directly to the nascent or misfolded polypeptide, frequently with concomitant ATP hydrolysis. The association of the chaperone prevents the polypeptide from aggregating with other polypeptides, and can force such aggregates to dissipate if they have already formed. After interaction with a second chaperone, GrpE (which permits an ADP-ATP exchange to occur), the polypeptide is released in a molten globule state and is permitted to fold. If misfolding occurs, the chaperones again associate with the misfolded protein, forcing it to return to an unfolded state. This cycle may be repeated until the protein is correctly folded. Unlike the first type of chaperones, which simply bind to the polypeptide, the second group (e.g. GroEL/ES) not only bind to the polypeptide, but also completely surround it such that it is protected from the surrounding environment. The GroEL/ES complex is composed of 2 stacked 14-member rings having a hydrophobic interior surface, and a 7-membered ring 'cap'. The polypeptide is drawn into the channel in the center of this complex in an ATP-dependent reaction where it is able to fold without interference from other polypeptides. Incorrectly folded proteins are not released from the complex.

An important step in protein folding is the creation of disulfide bonds. These bonds, either within a subunit or between subunits of a protein, are critical for protein stability. Disulfide bonds form readily in aqueous solution, and incorrect disulfide bond formation is difficult to reverse without the aid of a reducing environment. To assist in this process of correct disulfide bond formation, thiol-containing molecules, such as glutathione or thioredoxin, and their respective oxidation/reduction systems are found in the cytosol of most cells (Loferer, H., Hennecke, H. (1994) *Trends in Biochemical Sciences* 19(4): 169-171).

There are times, however, when folding of nascent polypeptide chains is not desirable, such as when these polypeptides are to be secreted. The folding process generally results in the hydrophobic regions of the protein being in the center of the protein, away from aqueous solution, and the hydrophilic regions being presented at the outer surfaces of the protein. This conformational arrangement, while creating greater stability for the protein, makes it difficult for the protein to be translocated across membranes, since the hydrophobic core of the membrane is inherently incompatible with the hydrophilic exterior of the protein. Thus, proteins synthesized by the cell which must be secreted to the exterior of the cell (e.g., cell surface enzymes and membrane receptors) or which must be inserted into the membrane itself (e.g., transporter proteins and channel proteins) are generally secreted or inserted prior to folding. The same chaperones which prevent aggregation of nascent polypeptide chains also prevent folding of polypeptides until they are disengaged. Thus, these proteins may 'escort' nascent polypeptide chains to an appropriate cellular location where they either are removed, thereby permitting folding, or they transfer the polypeptide to a transport system which will either secrete the polypeptide or aid its insertion into a membrane.

A specialized protein machinery has evolved that specifically detects, binds, transports, and processes proteins bearing specific prosequences (these sequences are later removed from the protein by cleavage). This machinery consists of a number of proteins which are collectively termed the sec (type II secretion) system (for review, see Gilbert, M. et al. (1995) *Critical Reviews in Biotechnology* 15(1): 13-39 and references therein; Freudl, R. (1992) *Journal of Biotechnology* 23(3): 231-240 and references therein; Neidhardt, F.C. et al. (1996) *E. coli* and *Salmonella* ASM Press: Washington, D.C., p. 967-978; Binet, R. et al. (1997) *Gene* 192(1): 7-11; and Rapoport, T.A. (1986) *Critical Reviews in Biochemistry* 20(1): 73-137, and references therein). The sec system is composed of chaperones (e.g., SecA and SecB), integral membrane proteins, also called translocases (e.g., SecY, SecE, and SecG), and signal peptidases (e.g., LepB). The nascent polypeptide having a prosequence directing secretion is bound by SecB, which delivers it to SecA at the inner surface of the cell membrane. Sec A binds to the

prosequence and, upon ATP hydrolysis, inserts into the membrane and forces a portion of the polypeptide through the membrane as well. The remainder of the polypeptide is guided through the membrane by a complex of translocases, such as SecY, SecE, and SecG. Finally, the signal peptidase cleaves off the prosequence and the polypeptide is free on the extracellular side of the membrane, where it spontaneously folds.

Sec-independent secretion mechanisms are also known. For example, the signal recognition particle-dependent pathway involves the binding of a signal recognition particle (SRP) protein to the nascent polypeptide as it is being synthesized, forcing the ribosome to stall. A receptor for SRP at the inner surface of the membrane then binds the ribosome-polypeptide-SRP complex. Hydrolysis of GTP provides the energy necessary to transfer the complex to the sec translocase complex, where the nascent polypeptide is guided across the membrane as it is synthesized by the ribosome. Other secretion mechanisms specific to only a few proteins are also known to exist.

III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as SES nucleic acid and protein molecules, which participate in *C. glutamicum* DNA repair or recombination, in the transposition or other rearrangement of *C. glutamicum* DNA, in *C. glutamicum* gene expression (e.g., the processes of transcription or translation), or in protein folding or protein secretion from this microorganism. In one embodiment, the SES molecules participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. In a preferred embodiment, the activity of the SES molecules of the present invention with regard to DNA repair or recombination, transposition of DNA, gene expression, protein folding or protein secretion has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the SES molecules of the invention are modulated in activity, such that the *C. glutamicum* cellular processes in which the SES molecules participate (e.g., DNA repair or recombination, transposition of DNA, gene expression, protein folding, or protein secretion) are also altered in activity, resulting either directly or indirectly in a modulation of the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

The language, "SES protein" or "SES polypeptide" includes proteins which participate in a number of cellular processes related to *C. glutamicum* genetic stability, gene expression, protein folding, or protein secretion. For example, an SES protein may be involved in *C. glutamicum* DNA repair or recombination mechanisms, in

rearrangements of *C. glutamicum* genetic material (such as those mediated by transposons), in transcription or translation of genes in this microorganism, in the mediation of *C. glutamicum* protein folding (such as the activity of chaperones) or in secretion of proteins from *C. glutamicum* cells (e.g., the sec system). Examples of SES proteins include those encoded by the SES genes set forth in Table 1 and Appendix A. The terms "SES gene" or "SES nucleic acid sequence" include nucleic acid sequences encoding an SES protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of SES genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound. The term "DNA repair" is art-recognized and includes cellular mechanisms whereby errors in DNA (due either to damage, such as, but not limited to, ultraviolet radiation, methylases, low-fidelity replication, or mutagens) are excised and corrected. The term "recombination" or "DNA recombination" is art-recognized and includes cellular mechanisms whereby extensive DNA damage affecting both strands of a DNA molecule is corrected by homologous recombination with another, undamaged copy of the DNA molecule within the same cell. Such repairs are generally low-fidelity, and may result in genetic rearrangements.

The term "transposon" is art-recognized and includes a DNA element which is able to insert randomly throughout the genome of an organism, and which may result in the disruption of genes or their regulatory regions, or in duplications, inversions, deletions, and other genetic rearrangements. The term "protein folding" is art-recognized and includes the movement of a polypeptide chain through multiple three-dimensional configurations until the stable, active, three-dimensional configuration is attained. The formation of disulfide bonds and the sequestration of hydrophobic regions from the surrounding aqueous solution provide some of the driving forces for this folding process, and correct folding may be enhanced by the activity of chaperones. The terms "secretion" or "protein secretion" is art-recognized and includes the movement of proteins from the interior of the cell to the exterior of the cell, in a mechanism whereby a system of secretion proteins permits their transit across the cellular membrane to the exterior of the cell.

In another embodiment, the SES molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an SES protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. For example, modulation of proteins involved directly in transcription or translation (e.g., polymerases or ribosomes) such that they are increased in number or in activity should increase global cellular transcription or translation (or rates of these processes). This increased cellular gene expression should include those proteins involved in fine chemical biosynthesis, so an increase in yield, production, or efficiency of production of one or more desired compounds may occur. Modifications to the transcriptional/translational protein machinery of *C. glutamicum* such that the regulation of these proteins is altered may also permit increased expression of genes involved in the production of fine chemicals. Modulation of the activity or number of proteins involved in polypeptide folding may permit an increase in the overall production of correctly folded molecules in the cell, thereby increasing the possibility that desired proteins (e.g., fine chemical biosynthetic proteins) are able to function properly. Further, by mutating proteins involved in secretion from *C. glutamicum* such that they are increased in number or activity, it may be possible to increase the secretion of a fine chemical (e.g., an enzyme) from cells in fermentor culture, where it may be readily recovered.

Genetic modification of the SES molecules of the invention may also result in indirect modulation of production of one or more fine chemicals. For example, by increasing the number or activity of a DNA repair or recombination protein of the

invention, one may increase the ability of the cell to detect and repair DNA damage. This should effectively increase the ability of the cell to maintain a mutated gene within its genome, thereby increasing the likelihood that a transgene engineered *into C. glutamicum* (e.g., encoding a protein which will increase biosynthesis of a fine chemical) will not be lost during culture of the microorganism. Conversely, by decreasing the number or activity of one or more DNA repair or recombination proteins, it may be possible to increase the genetic instability of the organism. Such manipulations should improve the ability of the organism to be modified by mutagenesis without the introduced mutation being corrected. The same holds true for proteins involved in transposition or rearrangement of genetic elements in *C. glutamicum* (e.g., transposons). By mutagenizing these proteins such that they are either increased or decreased in number or activity, it is possible to simultaneously increase or decrease the genetic stability of the microorganism. This has a profound impact on the ability of any other mutation to be introduced into *C. glutamicum*, and on the ability of introduced mutations to be retained. Transposons also offer a convenient mechanism by which mutagenesis of *C. glutamicum* may be performed; duplication of desired genes (e.g., fine chemical biosynthetic genes) is readily accomplished by transposon mutagenesis, as is disruption of undesired genes (e.g., genes encoding proteins involved in degradation of desired fine chemicals).

By modulating one or more proteins (e.g. sigma factors) involved in the regulation of transcription or translation in response to particular environmental conditions, it may be possible to prevent the cell from slowing or stopping protein synthesis under unfavorable environmental conditions, such as those found in large-scale fermentor culture. This should lead to increased gene expression, which in turn may permit increased biosynthesis of desired fine chemicals under such conditions. Many such secreted proteins have functions critical for cell viability (e.g., cell surface proteases or receptors). An alteration of a secretory pathway such that these proteins are more readily transported to their extracellular location may improve the overall viability of the cell, and thus result in greater numbers of *C. glutamicum* cells capable of producing fine chemicals during large-scale culture. Further, since certain bacterial protein secretion pathways (e.g., the sec system) are known to participate in the insertion of integral membrane proteins (such as receptors, channels, pores, or transporters) into the membrane, the modulation of activity of proteins involved in protein secretion from *C. glutamicum* may affect the ability of the cell to excrete waste products or to import necessary metabolites. If the activity of these secretory proteins is increased, then the ability of the cell to produce fine chemicals may be similarly increased (due to an increase in the presence of transporters/channels in the membrane which may import

nutrients or excrete waste products). If the activity of these proteins is decreased, then there may be insufficient nutrients available to support overproduction of desired compounds, or waste products may interfere with fine chemical biosynthesis.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C glutamicum* SES cDNAs and the predicted amino acid sequences of the *C. glutamicum* SES proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins involved in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The SES protein or a biologically active portion or fragment thereof of the invention can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections:

A. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode SES polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of SES-encoding nucleic acid (e.g., SES DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic

DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated SES nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* SES cDNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can

be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an SES nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* SES cDNAs of the invention. This cDNA comprises sequences encoding SES proteins (i.e., "the coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA number having the designation "RXA" followed by 5 digits (i.e., RXA00005). Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequence in Appendix B designated RXA00005 is a translation of the coding region of the nucleotide sequence of nucleic acid molecule RXA00005 in Appendix A.

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., et al. (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the

nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an SES protein. The nucleotide sequences determined from the cloning of the SES genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning SES homologues in other cell types and organisms, as well as SES homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone SES homologues. Probes based on the SES nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an SES protein, such as by measuring a level of an SES-encoding nucleic acid in a sample of cells, e.g., detecting SES mRNA levels or determining whether a genomic SES gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently

homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. Proteins involved in *C. glutamicum* genetic stability, gene expression, protein folding or protein secretion, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an SES protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of SES protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

Portions of proteins encoded by the SES nucleic acid molecules of the invention are preferably biologically active portions of one of the SES proteins. As used herein, the term "biologically active portion of an SES protein" is intended to include a portion, e.g., a domain/motif, of an SES protein that participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has an activity as set forth in Table 1. To determine whether an SES protein or a biologically active portion thereof can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an SES protein can be prepared by isolating a portion of one of the sequences in Appendix

B, expressing the encoded portion of the SES protein or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the SES protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same SES protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In addition to the *C. glutamicum* SES nucleotide sequences shown in Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of SES proteins may exist within a population (e.g., the *C. glutamicum* population). Such genetic polymorphism in the SES gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an SES protein, preferably a *C. glutamicum* SES protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the SES gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in SES that are the result of natural variation and that do not alter the functional activity of SES proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* SES cDNA of the invention can be isolated based on their homology to the *C. glutamicum* SES nucleic acid disclosed herein using the *C. glutamicum* cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even

more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* SES protein.

In addition to naturally-occurring variants of the SES sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded SES protein, without altering the functional ability of the SES protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the SES proteins (Appendix B) without altering the activity of said SES protein, whereas an "essential" amino acid residue is required for SES protein activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having SES activity) may not be essential for activity and thus are likely to be amenable to alteration without altering SES activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding SES proteins that contain changes in amino acid residues that are not essential for SES activity. Such SES proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the SES activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participating in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B,

more preferably at least about 60-70% homologous to one of the sequences in Appendix B. even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an SES protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an SES protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an SES coding sequence, such as by saturation mutagenesis, and the

resultant mutants can be screened for an SES activity described herein to identify mutants that retain SES activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding SES proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire SES coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an SES protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID RXA00005 comprises nucleotides 1 to 1608). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding SES. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding SES disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of SES mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of SES mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of SES mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense

nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an SES protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave SES mRNA transcripts to thereby inhibit translation of SES mRNA. A ribozyme having specificity for an SES-encoding nucleic acid can be designed based upon the nucleotide sequence of an SES cDNA disclosed herein (i.e., RXA00005 in Appendix A). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an SES-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742. Alternatively, SES mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, SES gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an SES nucleotide sequence (e.g., an SES promoter and/or enhancers) to form triple helical structures that prevent transcription of an SES gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

B Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an SES protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are

often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which are operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., SES proteins, mutant forms of SES proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of SES proteins in prokaryotic or eukaryotic cells. For example, SES genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. et al. (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. et al. (1991) "Heterologous gene expression in filamentous fungi" in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High

efficiency *Agrobacterium tumefaciens* –mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586). or mammalian cells. Suitable host cells are discussed further in Goeddel. *Gene Expression Technology: Methods in Enzymology* 185. Academic Press. San Diego. CA (1990). Alternatively. the recombinant expression vector can be transcribed and translated *in vitro*. for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein. usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often. in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes. and their cognate recognition sequences, include Factor Xa. thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith. D.B. and Johnson. K.S. (1988) *Gene* 67:31-40). pMAL (New England Biolabs. Beverly. MA) and pRIT5 (Pharmacia. Piscataway, NJ) which fuse glutathione S-transferase (GST). maltose E binding protein. or protein A. respectively. to the target recombinant protein. In one embodiment. the coding sequence of the SES protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising. from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant SES protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al.. (1988) *Gene* 69:301-315) and pET 11d (Studier et al.. *Gene Expression Technology: Methods in Enzymology* 185. Academic Press, San Diego. California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ

prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the SES protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

Alternatively, the SES proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the SES proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl. Acid. Res.* 12: 8711-8721.

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements.

For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to SES mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an SES protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning A Laboratory Manual, 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an SES protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an SES gene into which a deletion, addition or substitution

has been introduced to thereby alter, e.g., functionally disrupt, the SES gene. Preferably, this SES gene is a *Corynebacterium glutamicum* SES gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous SES gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous SES gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous SES protein). In the homologous recombination vector, the altered portion of the SES gene is flanked at its 5' and 3' ends by additional nucleic acid of the SES gene to allow for homologous recombination to occur between the exogenous SES gene carried by the vector and an endogenous SES gene in a microorganism. The additional flanking SES nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (e.g., by electroporation) and cells in which the introduced SES gene has homologously recombined with the endogenous SES gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an SES gene on a vector placing it under control of the lac operon permits expression of the SES gene only in the presence of IPTG. Such regulatory systems are well known in the art.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an SES protein. Accordingly, the invention further provides methods for producing SES proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an SES protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered SES protein) in a suitable medium until SES protein is produced. In another embodiment, the method further comprises isolating SES proteins from the medium or the host cell.

C. Isolated SES Proteins

Another aspect of the invention pertains to isolated SES proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of SES protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of SES protein having less than about 30% (by dry weight) of non-SES protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-SES protein, still more preferably less than about 10% of non-SES protein, and most preferably less than about 5% non-SES protein. When the SES protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of SES protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of SES protein having less than about 30% (by dry weight) of chemical precursors or non-SES chemicals, more preferably less than about 20% chemical precursors or non-SES chemicals, still more preferably less than about 10% chemical precursors or non-SES chemicals, and most preferably less than about 5% chemical precursors or non-SES chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the SES protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* SES protein in a microorganism such as *C. glutamicum*.

An isolated SES protein or a portion thereof of the invention can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the repair or recombination of DNA, in the transposition of genetic material, in gene

expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an SES protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the SES protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the SES protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences of Appendix B. The preferred SES proteins of the present invention also preferably possess at least one of the SES activities described herein. For example, a preferred SES protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the repair or recombination of DNA, in the transposition of genetic material, in gene expression (i.e., the processes of transcription or translation), in protein folding, or in protein secretion in *Corynebacterium glutamicum*, or which has one or more of the activities set forth in Table 1.

In other embodiments, the SES protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the SES protein is a protein which comprises an amino acid sequence which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the SES activities described herein. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an SES protein include peptides comprising amino acid sequences derived from the amino acid sequence of an SES protein, e.g., the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an SES protein, which include fewer amino acids than a full length SES protein or the full length protein which is homologous to an SES protein, and exhibit at least one activity of an SES protein. Typically, biologically active portions (peptides,

e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an SES protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an SES protein include one or more selected domains/motifs or portions thereof having biological activity.

SES proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the SES protein is expressed in the host cell. The SES protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an SES protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native SES protein can be isolated from cells (e.g., endothelial cells), for example using an anti-SES antibody, which can be produced by standard techniques utilizing an SES protein or fragment thereof of this invention.

The invention also provides SES chimeric or fusion proteins. As used herein, an SES "chimeric protein" or "fusion protein" comprises an SES polypeptide operatively linked to a non-SES polypeptide. An "SES polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an SES protein, whereas a "non-SES polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the SES protein, e.g., a protein which is different from the SES protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the SES polypeptide and the non-SES polypeptide are fused in-frame to each other. The non-SES polypeptide can be fused to the N-terminus or C-terminus of the SES polypeptide. For example, in one embodiment the fusion protein is a GST-SES fusion protein in which the SES sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant SES proteins. In another embodiment, the fusion protein is an SES protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of an SES protein can be increased through use of a heterologous signal sequence.

Preferably, an SES chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with

conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An SES-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the SES protein.

Homologues of the SES protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the SES protein. As used herein, the term "homologue" refers to a variant form of the SES protein which acts as an agonist or antagonist of the activity of the SES protein. An agonist of the SES protein can retain substantially the same, or a subset, of the biological activities of the SES protein. An antagonist of the SES protein can inhibit one or more of the activities of the naturally occurring form of the SES protein, by, for example, competitively binding to a downstream or upstream member of a biochemical cascade which includes the SES protein, by binding to a target molecule with which the SES protein interacts, such that no function interaction is possible, or by binding directly to the SES protein and inhibiting its normal activity.

In an alternative embodiment, homologues of the SES protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the SES protein for SES protein agonist or antagonist activity. In one embodiment, a variegated library of SES variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of SES variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential SES sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of SES sequences therein. There are a variety of methods which can be used to produce libraries of potential SES homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding

the desired set of potential SES sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the SES protein coding can be used to generate a variegated population of SES fragments for screening and subsequent selection of homologues of an SES protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an SES coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the SES protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of SES homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify SES homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated SES library, using methods well known in the art.

D. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of SES protein

regions required for function: modulation of an SES protein activity; modulation of the metabolism of one or more cell membrane components; modulation of the transmembrane transport of one or more compounds; and modulation of cellular production of a desired compound, such as a fine chemical.

The SES nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. Detection of such organisms is of significant clinical relevance.

Further, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The SES nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for

the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the SES nucleic acid molecules of the invention may result in the production of SES proteins having functional differences from the wild-type SES proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

The modulation of activity of proteins involved in *C. glutamicum* DNA repair, recombination, or transposition should impact the genetic stability of the cell. For example, by decreasing the number or activity of proteins involved in DNA repair mechanisms, one may decrease the ability of the cell to correct genetic errors, which should permit the simplified introduction of desired mutations into the genome (such as those encoding proteins involved in fine chemical production). Increasing the activity or number of transposons should result in a similarly increased mutation rate in the genome, and can permit facile duplication of desired genes (e.g., those encoding fine chemical biosynthetic proteins) or disruption of undesired genes (e.g., those encoding fine chemical degradation proteins). Conversely, by decreasing the number or activity of transposons or by increasing the number or activity of DNA repair proteins, it may be possible to increase the genetic stability of *C. glutamicum*, which in turn should result in better retention of introduced mutations in this microorganism through multiple generations in culture. Ideally, during mutagenesis and strain construction, one or more DNA repair systems would be decreased in activity and one or more transposons may be increased in activity, but once the desired mutation had been achieved in a strain, these the reverse would occur. Such manipulation is possible by placement of one or more DNA repair genes or transposons under control of an inducible repressor.

Modulation of proteins involved in transcription and translation in *C. glutamicum* can have both direct and indirect effects on the production of a fine chemical from these microorganisms. For example, by manipulating a protein which directly translates a gene (e.g., a polymerase) or which directly regulates transcription (e.g., a repressor or activator protein), it is possible to directly affect the expression of the target gene. In the case of genes encoding a protein involved in the biosynthesis or degradation of a fine chemical, this type of genetic manipulation should have a direct effect on the production of this fine chemical. Mutagenesis of a repressor protein such that it can no longer repress its target gene, or mutagenesis of an activator protein such that it is optimized in activity should lead to an increase in transcription of the target gene. If the target gene is, for example, a fine chemical biosynthetic gene, then an increase in production of that chemical may result, due to the overall greater number of

transcripts present for the gene, which should result in greater numbers of the protein as well. Increasing the number or activity of a repressor protein for a target sequence or decreasing the number or activity of an activator protein for a target sequence when this sequence is, for example, a fine chemical degradative protein, then a similar increase in production of the fine chemical should result.

Indirect effects on fine chemical production may also arise due to manipulation of proteins involved in transcription and translation. For example, by modulating the activity or number of transcription factors (e.g., the sigma factors) or translational repressors/activators which globally regulate transcription in *C. glutamicum* in response to environmental or metabolic factors, it should be possible to uncouple cellular transcription from environmental or metabolic regulation. In turn, this may permit continued transcription under conditions which would normally slow or altogether stop gene expression, such as those unfavorable conditions (e.g., high temperature, low oxygen, high waste product levels) which exist in large-scale fermentor cultures. By increasing the rate of gene (e.g., fine chemical biosynthetic gene) expression in such situations, the overall rate of fine product production may also be increased, at least due to the relatively greater number of fine chemical biosynthetic proteins in the cell. Principles and examples for modification of transcription and transcriptional regulation are described in, e.g., Lewin, B. (1990) Genes IV, Part 3: "Controlling procaryotic genes by transcription" Oxford Univ. Press: Oxford, p. 213-301.

Modulation of the activity or number of proteins involved in polypeptide folding (e.g., chaperones) may permit an increase in the overall production of correctly folded molecules in the cell. This has two effects: first, an overall increase in the number of proteins in the cell, due to the fact that fewer proteins are misfolded and degraded, and second, an increase in the number of any given protein that is correctly folded and thus active (see, e.g., Thomas, J.G., Baneyx, F. (1997) *Protein Expression and Purification* 11(3): 289-296; Luo, Z.H., and Hua, Z.C. (1998) *Biochemistry and Molecular Biology International* 46(3): 471-477; Dale, G.E., et al. (1994) *Protein Engineering* 7(7): 925-931; Amrein, K.E. et al. (1995) *Proc. Natl Acad. Sci. U.S.A.* 92(4): 1048-1052; and Caspers, P. et al. (1994) *Cell. Mol. Biol.* 40(5): 635-644). While such mutations result in an increase in the number of active proteins of all kinds, when coupled with additional mutations increasing the activity or number of, e.g., a fine chemical biosynthetic protein, an additive effect in the amount of correctly folded, active desired protein may be obtained.

Manipulation of proteins involved in secretion of polypeptides from *C. glutamicum* such that they are improved in activity or number may directly improve the secretion of a proteinaceous fine chemical (e.g., an enzyme) from this microorganism. It

is significantly easier to harvest and purify fine chemicals when they are secreted into the medium of large-scale cultures than when they are retained in the cell, so the yield and production of a fine chemical should be increased through such secretion system engineering. Genetic manipulation of these secretion proteins may also result in indirect improvements in the production of one or more fine chemicals. First, increased or decreased activity of one or more *C. glutamicum* secretion systems (as brought about by mutagenesis of one or more SES proteins involved in such pathways) may result in increased or decreased global secretion rates from the cell. Many such secreted proteins have functions critical for cell viability (e.g., cell surface proteases or receptors). An alteration of a secretory pathway such that these proteins are more readily transported to their extracellular location may improve the overall viability of the cell, and thus result in greater numbers of *C. glutamicum* cells capable of producing fine chemicals during large-scale culture. Second, certain bacterial secretion systems, (e.g., the sec system) are known to play a significant role in the process by which integral membrane proteins (e.g. channels, pores, or transporters) insert into cellular membranes. If the activity of one or more secretory pathway proteins is increased, then the ability of the cell to produce fine chemicals may be similarly increased, due to the presence of increased intracellular nutrient levels or decreased intracellular waste levels. If the activity of one or more such secretory pathway protein is decreased, then there may be insufficient nutrients available to support overproduction of desired compounds, or waste products may interfere with the biosynthesis of desired fine chemicals.

The aforementioned mutagenesis strategies for SES proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one skilled in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated SES nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any product produced by *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

Exemplification

Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC 13032

A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $MgSO_4 \times 7H_2O$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(NH_4)_2SO_4$, 1 g/l NaCl, 2 g/l $MgSO_4 \times 7H_2O$, 0.2 g/l $CaCl_2$, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $FeSO_4 \times H_2O$, 10 mg/l $ZnSO_4 \times 7\ H_2O$, 3 mg/l $MnCl_2 \times 4\ H_2O$, 30 mg/l H_3BO_3, 20 mg/l $CoCl_2 \times 6\ H_2O$, 1 mg/l $NiCl_2 \times 6\ H_2O$, 3 mg/l $Na_2MoO_4 \times 2\ H_2O$, 500 mg/l complexing agent (EDTA or critic acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca.18 h at 37°C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12.000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20

µg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

Starting from DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (*see* e.g., Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al* (1994) "Current Protocols in Molecular Biology". John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) Proc. Natl. Acad. Sci. USA, 75:3737-3741); pACYC177 (Change & Cohen (1978) J. Bacteriol 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) Gene 53:283-286.

Example 3: DNA Sequencing and Computational Functional Analysis

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R.D. et al. (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

Example 4: *In vivo* Mutagenesis

In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain

the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) Strategies 7: 32-34.

Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J.F. *et al.* (1987) Biotechnology, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E coli and C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. et al. (1985) J. Bacteriol. 162:591-597, Martin J.F. et al. (1987) Biotechnology, 5:137-146 and Eikmanns, B.J. et al. (1991) Gene, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C glutamicum* can be achieved by protoplast transformation (Kastsumata, R. et al. (1984) J. Bacteriol. 159306-311), electroporation (Liebl, E. et al. (1989) FEMS Microbiol. Letters, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A et al. (1990) J. Bacteriol. 172:1663-1666). It is also possible to transfer the shuttle vectors for

C. glutamicum to *E coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E coli* strain, such as NM522 (Gough & Murray (1983) J. Mol. Biol. 166:1-19).

Example 6: Assessment of the Expression of the Mutant Protein

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E.R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

Example 7: Growth of Genetically Modified *Corynebacterium glutamicum* — Media and Culture Conditions

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al.* (1989) Appl. Microbiol. Biotechnol., 32:205-210; von der Osten *et al.* (1998) Biotechnology Letters. 11:11-16; Patent DE 4,120.867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes. Volume II, Balows, A. *et al.*, eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources. inorganic salts. vitamins and trace elements. Preferred carbon sources are sugars, such as mono-. di-. or polysaccharides. For example, glucose, fructose. mannose. galactose, ribose. sorbose, ribulose. lactose. maltose. sucrose. raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts. such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates. urea. amino acids or complex nitrogen sources like corn steep liquor. soy bean flour. soy bean protein. yeast extract. meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium. magnesium, sodium, cobalt. molybdenum. potassium. manganese. zinc. copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids. such as citric acid. It is typical for the media to also contain other growth factors. such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin. folic acid. nicotinic acid. pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses. corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (*eds.* P.M. Rhodes, P.F. Stanbury. IRL Press (1997) pp. 53-73, ISBN 0

19 963577 3). It is also possible to select growth media from commercial suppliers. like standard 1 (Merck) or BHI (grain heart infusion. DIFC) or others.

All medium components are sterilized. either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or. if necessary. separately. All media components can be present at the beginning of growth. or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0. and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS. HEPES. ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced. due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the micro-organisms. the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels. such as microtiter plates. glass tubes. glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones. the microorganisms should be cultured in microtiter plates. glass tubes or shake flasks. either with or without baffles. Preferably 100 ml shake flasks are used. filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively. a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested. an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of 0.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose. 2.5 g/l NaCl, 2 g/l urea. 10 g/l polypeptone. 5 g/l yeast extract. 5 g/l meat extract. 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract. 5 g/l meat extract.

22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 8 – *In vitro* Analysis of the Function of Mutant Proteins

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes", VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) EMBO J. 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing

the medium and/or the cellular component for increased production of the desired product (i.e., an amino acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al. (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. et al. (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum*

cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biotekhnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Equivalents

Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

TABLE 1: GENES IN THE APPLICATION

Genes and enzymes involved in DNA uptake, repair and recombination

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA01020	GR00291	998	1744	BS-ung,EC-ung	URACIL-DNA GLYCOSYLASE (EC 3.2.2.-)
RXA00484	GR00119	21602	20568	EC-phrB	DEOXYRIBODIPYRIMIDINE PHOTOLYASE (EC 4.1.99.3)
RXA02476	GR00715	10514	9636	EC-mutY,BS-yfhQ	A/G-SPECIFIC ADENINE GLYCOSYLASE (EC 3.2.2.-)
RXA00102	GR00014	11288	10521		FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (EC 3.2.2.23)
RXA01670	GR00466	3	614		FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (EC 3.2.2.23)
RXA02078	GR00628	8170	9027	EC-mutM,BS-mutM	FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (EC 3.2.2.23)
RXA00484	GR00119	21602	20568	EC-phrB	DEOXYRIBODIPYRIMIDINE PHOTOLYASE (EC 4.1.99.3)
RXA01596	GR00447	4370	6148	BS-recN,EC-recN	DNA REPAIR PROTEIN RECN
RXA01493	GR00423	7530	6220		DNA-DAMAGE-INDUCIBLE PROTEIN F
RXA02671	GR00753	11718	12296	BS-sms,EC-sms	DNA REPAIR PROTEIN RADA HOMOLOG
RXA02291	GR00662	1518	865	EC-alkB	DNA repair gene specific for alkylated DNA
RXA01733	GR00492	2	544		RECF PROTEIN
RXA01252	GR00365	643	1296	BS-recR,EC-recR	RECOMBINATION PROTEIN RECR
RXA01878	GR00537	1239	2117	BS-ksgA,EC-ksgA	DIMETHYLADENOSINE TRANSFERASE (EC 2.1.1.-)
RXA01556	GR00433	1	849		METHYLPHOSPHOTRIESTER-DNA ALKYLTRANSFERASE
RXA00053	GR00008	8162	8554	EC-b1759	MUTATOR MUTT PROTEIN (7,8-DIHYDRO-8-OXOGUANINE-TRIPHOSPHATASE) (8-OXO-DGTPASE) (EC 3.6.1.-)
RXA00280	GR00043	4198	4696		MUTATOR MUTT PROTEIN (7,8-DIHYDRO-8-OXOGUANINE-TRIPHOSPHATASE) (8-OXO-DGTPASE) (EC 3.6.1.-)
RXA00333	GR00057	16166	16699		MUTATOR MUTT PROTEIN (7,8-DIHYDRO-8-OXOGUANINE-TRIPHOSPHATASE) (8-OXO-DGTPASE) (EC 3.6.1.-)
RXA02110	GR00632	3641	4258		MUTATOR MUTT PROTEIN (7,8-DIHYDRO-8-OXOGUANINE-TRIPHOSPHATASE) (8-OXO-DGTPASE) (EC 3.6.1.-)
RXA02290	GR00662	693	295		DNA-3-METHYLADENINE GLYCOSIDASE I (EC 3.2.2.20)
RXA02557	GR00731	3766	3179	EC-tag	DNA-3-METHYLADENINE GLYCOSIDASE I (EC 3.2.2.20)
RXA02130	GR00638	1	87		DNA REPAIR HELICASE RAD25
RXA02131	GR00638	412	732		DNA REPAIR HELICASE RAD25
RXA02188	GR00641	23500	23967		DNA REPAIR HELICASE RAD25
RXA02742	GR00763	12384	10036		Hypothetical DNA Repair Helicase
RXA01979	GR00573	1	849		ATP-DEPENDENT DNA HELICASE RECQ (EC 3.6.1.-)
RXA02445	GR00709	9362	11050		ATP-DEPENDENT DNA HELICASE RECG
RXA00927	GR00253	1606	518	EC-ruvB,BS-ruvB	HOLLIDAY JUNCTION DNA HELICASE RUVB
RXA00928	GR00253	2233	1616	EC-ruvA,BS-ruvA	HOLLIDAY JUNCTION DNA HELICASE RUVA
RXA00172	GR00027	455	6	EC-pin	RESOLVASE
RXA00184	GR00028	8239	9411		DNA repair exonuclease
RXA00019	GR00002	14399	16258		SINGLE-STRANDED-DNA-SPECIFIC EXONUCLEASE RECJ (EC 3.1.-.-)
RXA00929	GR00253	2938	2276		CROSSOVER JUNCTION ENDODEOXYRIBONUCLEASE RUVC (EC 3.1.22.4)
RXA02251	GR00654	18367	18666		EXCINUCLEASE ABC SUBUNIT C
RXA02252	GR00654	18632	20455	BS-uvrC	EXCINUCLEASE ABC SUBUNIT C
RXA02416	GR00705	3	2642	BS-uvrA,EC-uvrA	EXCINUCLEASE ABC SUBUNIT A
RXA02563	GR00732	1515	2246		Excinuclease ATPase subunit
RXA02731	GR00762	3263	5359	BS-uvrB,EC-uvrB	EXCINUCLEASE ABC SUBUNIT B
RXA00531	GR00136	7854	7411	EC-dps	

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00998	GR00283	2871	2410		COMA OPERON PROTEIN 2
RXA02386	GR00693	1180	776	BS-comEA	COME OPERON PROTEIN 1, DNA binding and uptake (competence)
RXA02385	GR00693	776	6		COME OPERON PROTEIN 3, DNA binding and uptake (competence)
RXA02388	GR00694	1770	925		COME OPERON PROTEIN 3, DNA binding and uptake (competence)
RXA01975	GR00571	242	2137		PUTATIVE TYPE II RESTRICTION ENDONUCLEASE AND PUTATIVE TYPE I OR TYPE III RESTRICTION ENDONUCLEASE GENES, COMPLETE CDS
RXA01954	GR00562	3326	4165		TYPE III RESTRICTION-MODIFICATION SYSTEM ECOP15I ENZYME MOD (EC 2.1.1.72)
RXA02236	GR00654	4249	4566		integration host factor

Transposon, IS elements, Transposase, Integrase

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA02890	GR10027	112	1194		INTEGRASE
RXA01601	GR00447	11128	12039	BS-ripX,EC-xerD	INTEGRASE/RECOMBINASE XERD
RXA01228	GR00355	1668	1883		TRANSPOSONS TN1721 AND TN4653 RESOLVASE
RXA00263	GR00040	2243	936		DNA, TRANSPOSABLE ELEMENT IS31831
RXA01541	GR00428	3865	3095		PLASMID PASU1 TRANSPOSASE
RXA02590	GR00741	14837	13902		INSERTION ELEMENT IS1415 TRANSPOSASE I(ISTA) AND HELPER PROTEIN (ISTB) GENES, COMPLETE CDS
RXA00016	GR00002	8857	7964	EC-tra5_2	IS3 RELATED INSERTION ELEMENT
RXA00265	GR00040	2840	3289	BS-ydfG	TRANSPOSASE
RXA00938	GR00256	670	927		TRANSPOSASE
RXA01264	GR00367	12003	11788		TRANSPOSASE
RXA01265	GR00367	12616	12467		TRANSPOSASE
RXA01327	GR00386	753	898		TRANSPOSASE
RXA01328	GR00386	991	1365		TRANSPOSASE
RXA01329	GR00386	1407	1697		TRANSPOSASE
RXA01443	GR00418	13570	12740		TRANSPOSASE
RXA01444	GR00418	13928	13662		TRANSPOSASE
RXA01648	GR00457	829	461		TRANSPOSASE
RXA01649	GR00457	1260	841		TRANSPOSASE
RXA01650	GR00457	1437	1324		TRANSPOSASE
RXA01651	GR00457	1618	1484		TRANSPOSASE
RXA01680	GR00467	9590	9180		TRANSPOSASE
RXA01784	GR00505	3	551		TRANSPOSASE
RXA01662	GR00529	4961	6166		TRANSPOSASE
RXA01953	GR00562	928	548		TRANSPOSASE
RXA01998	GR00589	1345	2052		TRANSPOSASE
RXA02837	GR00829	179	6		

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00005	GR00001	4724	6331	EC-tra8_1	TRANSPOSASE
RXA00017	GR00002	9150	8857		TRANSPOSASE
RXA00057	GR00009	2491	2393		TRANSPOSASE
RXA00227	GR00032	27991	27194		TRANSPOSASE
RXA01819	GR00515	8287	7841		transposase

Aminoacyl-tRNA synthetases / tRNAs and tRNA metabolism

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02788	GR00777	2359	5022	BS-alaS,EC-alaS	ALANYL-TRNA SYNTHETASE (EC 6.1.1.7)
RXA00975	GR00275	780	4	BS-argS	POSSIBLE ARGINYL-TRNA SYNTHETASE (EC 6 1 1 19)
RXA00976	GR00275	1423	824		POSSIBLE ARGINYL-TRNA SYNTHETASE (EC 6 1 1 19)
RXA01730	GR00490	298	1974	BS-aspS,EC-aspS	ASPARTYL-TRNA SYNTHETASE (EC 6 1 1 12)
RXA00314	GR00053	5406	4027	EC-cysS,BS-cysS	CYSTEINYL-TRNA SYNTHETASE (EC 6 1.1.16)
RXA02204	GR00646	8756	7497		CYSTEINYL-TRNA SYNTHETASE (EC 6 1.1.16)
RXA01124	GR00312	2	1510	BS-glX,EC-glX	GLUTAMYL-TRNA SYNTHETASE (EC 6 1 1.17)
RXA00458	GR00115	232	5		GLUTAMYL-TRNA SYNTHETASE (EC 6 1 1.17)
RXA00069	GR00011	2782	1400		GLYCYL-TRNA SYNTHETASE (EC 6.1.1.14)
RXA01852	GR00525	4873	3587	BS-hisS,EC-hisS	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21)
RXA02726	GR00760	4530	1597	BS-ileS,EC-ileS	ISOLEUCYL-TRNA SYNTHETASE (EC 6.1 1 5)
RXA00966	GR00271	533	8		LEUCYL-TRNA SYNTHETASE (EC 6 1 1 4)
RXA01864	GR00531	474	4		LEUCYL-TRNA SYNTHETASE (EC 6 1 1 4)
RXA00968	GR00272	1007	6	BS-leuS,EC-leuS	LEUCYL-TRNA SYNTHETASE (EC 6 1 1 4)
RXA01061	GR00296	10974	10567		LEUCYL-TRNA SYNTHETASE (EC 6 1 1 4)
RXA01522	GR00424	26014	27591	BS-lysS,EC-lysU	LYSYL-TRNA SYNTHETASE (EC 6.1.1.6)
RXA02015	GR00609	152	670		METHIONYL-TRNA SYNTHETASE (EC 6 1 1.10)
RXA01582	GR00440	1619	2707	BS-pheS,EC-pheS	PHENYLALANYL-TRNA SYNTHETASE ALPHA CHAIN (EC 6 1 1.20)
RXA01583	GR00440	2914	4629	EC-pheT,BS-pheT	PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6 1 1.20)
RXA01717	GR00487	1000	719		PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6 1 1 20)
RXA01938	GR00556	94	1008		PROLYL-TRNA SYNTHETASE (EC 6.1.1.15)
RXA02692	GR00754	15485	16750	BS-serS,EC-serS	SERYL-TRNA SYNTHETASE (EC 6.1.1.11)
RXA02167	GR00640	13255	14514	BS-tyrS,EC-tyrS	TYROSYL-TRNA SYNTHETASE 1 (EC 6.1 1 1)
RXA02509	GR00721	2	1972	BS-thrZ,EC-thrS	THREONYL-TRNA SYNTHETASE (EC 6.1 1 3)
RXA02860	GR10006	2	439		TRYPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2)
RXA02866	GR10007	3992	4471	EC-trpS,BS-trpS	TRYPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2)
RXA00985	GR00279	498	4		VALYL-TRNA SYNTHETASE (EC 6.1 1 9)
RXA01347	GR00391	3038	5084	BS-valS,EC-valS	VALYL-TRNA SYNTHETASE (EC 6.1 1 9)
RXA00454	GR00112	869	6	BS-tgt,EC-tgt	QUEUINE TRNA-RIBOSYLTRANSFERASE (EC 2.4.2.29)

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01490	GR00423	3442	4332	BS-truB,EC-truB	TRNA PSEUDOURIDINE 55 SYNTHASE
RXA01621	GR00452	473	1912	BS-papS,EC-cca	TRNA NUCLEOTIDYLTRANSFERASE (EC 2.7.7.25)
RXA01704	GR00480	3	818		TRNA (URACIL-5-)-METHYLTRANSFERASE (EC 2.1.1.35)
RXA02523	GR00725	1587	2405	BS-trmD,EC-trmD	TRNA (GUANINE-N1)-METHYLTRANSFERASE (EC 2.1.1.31)
RXA02243	GR00654	11114	12058	EC-fmt,BS-fmt	METHIONYL-TRNA FORMYLTRANSFERASE (EC 2.1.2.9)
RXA00217	GR00032	17389	16295	EC-ycfB,BS-yrrA	PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE (EC 2.1.1.61)
RXA01223	GR00354	4156	3545	BS-spoVC,EC-pth	PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)
RXA01226	GR00354	7416	6973		PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29)
RXA00209	GR00032	9592	8102	BS-yerM	L-glutamyl-tRNA('Gln)-dependent amidotransferase subunit A (EC 6.3.5.-)
RXA00210	GR00032	9897	9601		L-glutamyl-tRNA('Gln)-dependent amidotransferase subunit C (EC 6.3.5.-)
RXA02686	GR00754	11266	10130		L-glutamyl-tRNA('Gln)-dependent amidotransferase subunit A (EC 6.3.5.-)
RXA02625	GR00747	791	6	BS-yerN	L-glutamyl-tRNA('Gln)-dependent amidotransferase subunit B (EC 6.3.5.-)
RXA01398	GR00408	7645	7010		L-glutamyl-tRNA('Gln)-dependent amidotransferase subunit B (EC 6.3.5.-)
RXA02228	GR00653	1876	2778	EC-miaA,BS-miaA	TRNA DELTA(2)-ISOPENTENYLPYROPHOSPHATE TRANSFERASE (EC 2.5.1.8)
RXA02502	GR00720	15510	16901	BS-hemA,EC-hemA	GLUTAMYL-TRNA REDUCTASE (EC 1.2.1.-)
RXA02182	GR00641	17875	18648		GLUTAMINE CYCLOTRANSFERASE PRECURSOR (EC 2.3.2.5), Glutaminyl-tRNA cyclotransferase

Transcription

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA01344	GR00390	2551	5	BS-rpoB,EC-rpoB	DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6)
RXA01387	GR00407	372	4		DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6)
RXA01388	GR00407	590	459		DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6)
RXA01389	GR00407	1350	667		DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6)
RXA01283	GR00369	7109	5817	BS-rpoC,EC-rpoC	DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6)
RXA01433	GR00417	9606	9004		SIGMA FACTOR
RXA02456	GR00712	1127	510		RNA POLYMERASE SIGMA-H FACTOR
RXA00304	GR00051	696	4		RNA POLYMERASE SIGMA FACTOR
RXA00485	GR00123	1210	1773		PUTATIVE RNA POLYMERASE SIGMA FACTOR CY78.15
RXA00532	GR00137	3	587		PROBABLE RNA POLYMERASE SIGMA FACTOR CY49.08
RXA01530	GR00426	1724	1083	BS-sigA,EC-rpoD	RNA POLYMERASE SIGMA FACTOR RPOD
RXA01531	GR00426	2565	1549		RNA POLYMERASE SIGMA FACTOR RPOD
RXA02065	GR00626	5348	5995	BS-sigW	EXTRACYTOPLASMIC FUNCTION ALTERNATIVE SIGMA FACTOR
RXA00588	GR00156	13672	14193		TRANSCRIPTION ELONGATION FACTOR GREA
RXA01723	GR00488	6500	7436		TRANSCRIPTION TERMINATION FACTOR RHO
RXA01724	GR00488	7429	7812		TRANSCRIPTION TERMINATION FACTOR RHO

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01725	GR00488	7897	8004		TRANSCRIPTION TERMINATION FACTOR RHO
RXA01726	GR00488	8000	8572	EC-rho,BS-rho	TRANSCRIPTION TERMINATION FACTOR RHO
RXA00736	GR00199	1	1887	BS-mfd,EC-mfd	TRANSCRIPTION-REPAIR COUPLING FACTOR
RXA00737	GR00200	1	480		TRANSCRIPTION-REPAIR COUPLING FACTOR
RXA01872	GR00535	768	4	BS-yybE,EC-cynR	TRANSCRIPTIONAL REGULATORY PROTEIN
RXA01081	GR00302	75	824		SPORULATION TRANSCRIPTION FACTOR
RXA02656	GR00752	10955	10116	BS-yojE,EC-raiD	chloramphenicol-sensitive proteinC
RXA02413	GR00703	3029	2538		PAPX PROTEIN, transcriptional regulator

Translation

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02418	GR00705	5101	5667	EC-infC,BS-infC	Bacterial Protein Translation Initiation Factor 3 (IF-3)
RXA00755	GR00203	1280	6		Protein Translation Initiation Factor 2 (IF-2)
RXA01496	GR00423	10908	9181	BS-infB,EC-infB	Protein Translation Initiation Factor 2 (IF-2)
RXA00677	GR00178	1624	1839	BS-infA,EC-infA	Bacterial Protein Translation Initiation Factor 1 (IF-1)
RXA01284	GR00370	510	4	BS-tufA,EC-tufB	Bacterial Protein Translation Elongation Factor Tu (EF-TU)
RXA00138	GR00022	1914	2474	BS-efp,EC-efp	Protein Translation Elongation Factor P (EF-P)
RXA01278	GR00369	2425	299	BS-fus,EC-fusA	Protein Translation Elongation Factor G (EF-G)
RXA01913	GR00547	1856	2680	BS-tsf	Protein translation Elongation Factor Ts (EF-Ts)
RXA00331	GR00057	15141	14785		Hypothetical Translational Inhibitor Protein
RXA02822	GR00803	1	570	BS-prfA,EC-prfA	Bacterial Peptide Chain Release Factor 1 (RF-1)
RXA00011	GR00002	2739	2383		Bacterial Peptide Chain Release Factor 2 (RF-2)
RXA00012	GR00002	3487	2612	EC-prfB	Bacterial Peptide Chain Release Factor 2 (RF-2)
RXA01926	GR00554	1	672		PEPTIDE CHAIN RELEASE FACTOR 3
RXA02002	GR00592	383	6		PEPTIDE CHAIN RELEASE FACTOR 3
RXA00896	GR00244	2884	3522		POLYPEPTIDE DEFORMYLASE (EC 3.5.1.31)
RXA02242	GR00654	10585	11091	EC-def,BS-def	POLYPEPTIDE DEFORMYLASE (EC 3.5.1.31)

Protein translocation, secretion, and folding

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01710	GR00484	850	443	EC-b1778,BS-yppQ	PEPTIDE METHIONINE SULFOXIDE REDUCTASE
RXA00124	GR00020	737	6		PREPROTEIN TRANSLOCASE SECA SUBUNIT
RXA00125	GR00020	1467	703		PREPROTEIN TRANSLOCASE SECA SUBUNIT

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00687	GR00179	9121	10440	EC-prlA,BS-secY	PREPROTEIN TRANSLOCASE SECY SUBUNIT
RXA00935	GR00254	654	4		PROTEIN-EXPORT MEMBRANE PROTEIN SECD
RXA01558	GR00434	1735	527		PROTEIN-EXPORT MEMBRANE PROTEIN SECF
RXA01559	GR00434	1983	1741		PROTEIN-EXPORT MEMBRANE PROTEIN SECD
RXA02260	GR00654	30280	30510		PROTEIN-EXPORT MEMBRANE PROTEIN SECG HOMOLOG
RXA02429	GR00707	4823	7111	BS-secA,EC-secA	PREPROTEIN TRANSLOCASE SECA SUBUNIT
RXA02462	GR00712	7653	6739		PREPROTEIN TRANSLOCASE SECA SUBUNIT
RXA02748	GR00764	2434	4074	BS-ffh,EC-ffh	SIGNAL RECOGNITION PARTICLE PROTEIN
RXA01355	GR00393	2877	3662		SIGNAL PEPTIDASE I (EC 3 4 21.89)
RXA00046	GR00007	5363	6058	EC-b3223	Signal recognition particle GTPase
RXA00758	GR00203	4797	6398		PERIPLASMIC OLIGOPEPTIDE-BINDING PROTEIN PRECURSOR
RXA01346	GR00391	1409	2983		PERIPLASMIC OLIGOPEPTIDE-BINDING PROTEIN PRECURSOR
RXA00453	GR00111	2393	45	BS-ydiJ	Preprotein translocase subunit
RXA00753	GR00202	23301	21880		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA01179	GR00335	4639	5151		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA01274	GR00367	27148	28242		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA01449	GR00419	1046	6		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA01798	GR00509	276	4		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA01818	GR00515	6453	7439		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA02607	GR00742	13971	14189		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA02608	GR00742	14248	15942		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA02886	GR10021	1907	2737		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA02894	GR10036	1017	232		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA02904	GR10042	686	12		PS1 PROTEIN PRECURSOR (PS1, one of the two major secreted proteins of Corynebacterium glutamicum)
RXA02025	GR00614	862	212	EC-msrA,BS-yppP	PEPTIDE METHIONINE SULFOXIDE REDUCTASE
RXA00107	GR00014	17940	18176		GLUTAREDOXIN-LIKE PROTEIN NRDH
RXA01431	GR00417	7858	7538	BS-trxA,EC-trxA	THIOREDOXIN REDUCTASE (EC 1 6 4 5) / THIOREDOXIN
RXA01432	GR00417	8896	7946	EC-trxB,BS-trxB	THIOREDOXIN REDUCTASE (EC 1 6 4 5)
RXA00937	GR00256	1	123		THIOREDOXIN
RXA01199	GR00343	3813	4583		THIOREDOXIN
RXA01613	GR00449	7055	5841	EC-gor	GLUTATHIONE REDUCTASE (EC 1 6 4 2)
RXA00539	GR00139	1460	1936	BS-bsaA,EC-bluE	GLUTATHIONE PEROXIDASE (EC 1.11 1.9)
RXA00824	GR00221	4356	4913		THIOL-DISULFIDE INTERCHANGE PROTEIN TLPA
RXA01841	GR00522	115	477		THIOL DISULFIDE INTERCHANGE PROTEIN TLPA
RXA01863	GR00530	830	24	BS-yfkB	/O/C Thioredoxin-like oxidoreductases
RXA02323	GR00668	1429	506		THIOREDOXIN REDUCTASE (EC 1 6 4 5)
RXA01072	GR00300	377	147	EC-b2673	NRDH-REDOXIN
RXA02436	GR00709	1596	1036		PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC 5.2.1.8)
RXA01837	GR00518	858	466		PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC 5.2 1 8)
RXA02047	GR00624	1	192		PROLYL ENDOPEPTIDASE (EC 3.4.21.26)
RXA02174	GR00641	9290	8937		PROBABLE FK506-BINDING PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PPIASE) (EC 5.2 1 8)
RXA00568	GR00152	2926	1582	BS-tig,EC-tig	TRIGGER FACTOR

TABLE 2: GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al "DNA fragment coding for phosphoenolpyruvat corboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 03/21/90
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moeckel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent. WO 9519442-A 5 07/20/95
AB003132	murC, ftsQ; ftsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from coryneform bacteria," *Biochem. Biophys. Res. Commun.*, 236(2):383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from Coryneform bacteria," *Appl. Microbiol Biotechnol*, 51(2):223-228 (1999)
AB018530	dtsR		Kimura, E. et al. "Molecular cloning of a novel gene, dtsR, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*," *Biosci. Biotechnol. Biochem*, 60(10):1565-1570 (1996)
AB018531	dtsR1; dtsR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	gltB; gltD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamolytransferase	
AF036932	aroD	3-dehydroquinate dehydratase	

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF038548	pyc	Pyruvate carboxylase	
AF038651	dciAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	Wehmeier, L. et al. "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp metabolism," *Microbiology*, 144.1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase, ornithine acetyltransferase; N-acetylglutamate kinase, acetylornithine transminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in Corynebacterium glutamicum," *Mol Cells.*, 8(3):286-294 (1998)
AF053071	aroB	Dehydroquinate synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Dusch, N. et al. "Expression of the Corynebacterium glutamicum panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in Escherichia coli," *Appl. Environ. Microbiol.*, 65(4)1530-1539 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pcpQ	Chorismate synthase; shikimate kinase; 3-dehydroquinate synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete')	Wehrmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with Corynebacterium glutamicum," *J Bacteriol.*, 180(12):3159-3165 (1998)
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase; ?; high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD, srp; amtP	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzmye); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in Corynebacterium glutamicum; Isolation of genes involved in biochemical characterization of corresponding proteins," *FEMS Microbiol*, 173(2):303-310 (1999)
AJ132968	cat	Chloramphenicol aceteyl transferase	
AJ224946	mqo	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from Corynebacterium glutamicum," *Eur J Biochem*, 254(2):395-403 (1998)
AJ238250	ndh	NADH dehydrogenase	
AJ238703	porA	Porin	Lichtinger, T. et al "Biochemical and biophysical characterization of the cell wall porin of Corynebacterium glutamicum: The channel is formed by a low molecular mass polypeptide," *Biochemistry*, 37(43):15024-15032 (1998)
D17429		Transposable element IS31831	Vertes, A.A. et al. "Isolation and characterization of IS31831, a transposable element from Corynebacterium glutamicum," *Mol Microbiol*, 11(4):739-746 (1994)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the Corynebacterium glutamicum (Brevibacterium lactofermentum AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," *Microbiology*, 142.3347-3354 (1996)
E01358	hdh, hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 1 10/12/87
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 2 10/12/87
E01375		Tryptophan operon	
E01376	trpL; trpE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent. JP 1987244382-A 1 10/24/87
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E03937		Biotin-synthase	Hatakeyama, K. et al "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 10/02/92
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04307		Flavum aspartase	Kurusu, Y. et al "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 02/09/93
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R et al. "Gene manifestation controlling DNA," Patent. JP 1993056782-A 3 03/09/93
E04484		Prephenate dehydratase	Sotouchi, N. et al "Production of L-phenylalanine by fermentation," Patent. JP 1993076352-A 2 03/30/93
E05108		Aspartokinase	Fugono, N et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 07/27/93
E05112		Dihydro-dipichorinate synthetase	Hatakeyama, K. et al "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 07/27/93

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent. JP 1993284970-A 1 11/02/93
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent. JP 1993284972-A 1 11/02/93
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent. JP 1993344881-A 1 12/27/93
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent JP 1993344893-A 1 12/27/93
E06825		Aspartokinase	Sugimoto, M. et al "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membraneous protein to membrane," Patent: JP 1994169780-A 1 06/21/94
E08177		Aspartokinase	Sato, Y et al "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08178, E08179, E08180, E08181, E08182		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08232		Acetohydroxy-acid isomeroreductase	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroreductase," Patent. JP 1994277067-A 1 10/04/94
E08234	secE		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 10/04/94
E08643		FT aminotransferase and desthiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08649		Aspartase	Kohama, K. et al "DNA fragment having promoter function in coryneform bacterium," Patent. JP 1995031478-A 1 02/03/95
E08900		Dihydrodipicolinate reductase	Madori, M. et al "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 03/20/95
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent. JP 1995075579-A 1 03/20/95
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-trypophan," Patent· JP 1997028391-A 1 02/04/97
E12760, E12759, E12758		transposase	Moriya, M. et al "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12764		Arginyl-tRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12767		Dihydrodipicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12773		Dihydrodipicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent· JP 1997070291-A 03/18/97
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent· JP 1997224661-A 1 09/02/97
L01508	IlvA	Threonine dehydratase	Moeckel, B. et al. "Functional and structural analysis of the threonine dehydratase of Corynebacterium glutamicum," *J. Bacteriol*, 174.8065-8072 (1992)
L07603	EC 4.2 1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C et al. "The cloning and nucleotide sequence of Corynebacterium glutamicum 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," *FEMS Microbiol Lett.*, 107:223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit; Acetohydroxy acid isomeroreductase	Keilhauer, C. et al. "Isoleucine synthesis in Corynebacterium glutamicum. molecular analysis of the ilvB-ilvN-ilvC operon," *J Bacteriol.*, 175(17).5595-5603 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
L18874	PtsM	Phosphoenolpyruvate sugar phosphotransferase	Fouet, A et al "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system: expression in Escherichia coli and homology to enzymes II from enteric bacteria," *PNAS USA*, 84(24) 8773-8777 (1987); Lee, J.K. et al. "Nucleotide sequence of the gene encoding the Corynebacterium glutamicum mannose enzyme II and analyses of the deduced protein sequence," *FEMS Microbiol Lett*, 119(1-2).137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium glutamicum," *J Microbiol. Biotechnol*, 4(4) 256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum," *Appl. Environ Microbiol.*, 60(7):2501-2507 (1994)
L28760	aceA	Isocitrate lyase	
L35906	dtxt	Diphtheria toxin repressor	Oguiza, J.A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the Corynebacterium diphtheriae dtxR from Brevibacterium lactofermentum," *J. Bacteriol*, 177(2):465-467 (1995)
M13774		Prephenate dehydratase	Follettie, M.T. et al. "Molecular cloning and nucleotide sequence of the Corynebacterium glutamicum pheA gene," *J Bacteriol.*, 167:695-702 (1986)
M16175	5S rRNA		Park, Y-H et al. "Phylogenetic analysis of the coryneform bacteria by 56 rRNA sequences," *J Bacteriol*, 169:1801-1806 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52.191-200 (1987)
M16664	trpA	Tryptophan synthase, 3' end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52:191-200 (1987)
M25819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of Corynebacterium glutamicum ATCC13032," *Gene*, 77(2):237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen Microbiol*, 138.1167-1175 (1992)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
M85107, M85108		23S rRNA gene insertion sequence	Rollet, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen Microbiol*., 138:1167-1175 (1992)
M89931	aecD; bmQ, yhbw	Beta C-S lyase, branched-chain amino acid uptake carrier, hypothetical protein yhbw	Rossol, I. et al "The Corynebacterium glutamicum aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," *J. Bacteriol.*, 174(9):2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the bmQ gene product," *Arch Microbiol.*, 169(4):303-312 (1998)
S59299	trp	Leader gene (promoter)	Herry, D M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: identification of a mutation in the trp leader sequence," *Appl. Environ. Microbiol*., 59(3):791-799 (1993)
U11545	trpD	Anthranilate phosphoribosyltransferase	O'Gara, J.P. and Dunican, L.K (1994) Complete nucleotide sequence of the Corynebacterium glutamicum ATCC 21850 tpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR, clglIR	Putative type II 5-cytosoine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from Corynebacterium glutamicum ATCC 13032 and analysis of its role in intergeneric conjugation with Escherichia coli," *J Bacteriol.*, 176(23):7309-7319 (1994); Schafer, A. et al. "The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine in an McrBC-deficient Escherichia coli strain," *Gene*, 203(2):95-101 (1997)
U14965	recA		
U31224	ppx		Ankri, S et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J Bacteriol.*, 178(15):4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J Bacteriol*, 178(15):4412-4419 (1996)
U31230	obg; proB, unkdh	?;gamma glutamyl kinase;similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway. A natural bypass of the proA step," *J Bacteriol*, 178(15).4412-4419 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
U31281	bioB	Biotin synthase	Serebriiskii, I.G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of Methylobacillus flagellatum and Corynebacterium glutamicum," *Gene*, 175:15-22 (1996)
U35023	thiR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A Corynebacterium glutamicum gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," *Arch Microbiol.*, 166(2):76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A Corynebacterium glutamicum gene conferring multidrug resistance in the heterologous host Escherichia coli," *J Bacteriol.*, 179(7):2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'5''-aminoglycoside phosphotransferase	
U89648		Corynebacterium glutamicum unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the Brevibacterium lactofermentum tryptophan operon," *Nucleic Acids Res.*, 14(24):10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of Corynebacterium glutamicum and possible mechanisms for modulation of its expression," *Mol Gen. Genet.*, 212(1):112-119 (1988)
X14234	EC 4 1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B.J. et al. "The Phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: Molecular cloning, nucleotide sequence, and expression," *Mol Gen. Genet.*, 218(2):330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," *Plant Mol. Biol.*, 21 (3):487-502 (1993)
X17313	fda	Fructose-bisphosphate aldolase	Von der Osten, C.H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the Corynebacterium glutamicum fda gene: structural comparison of C. glutamicum fructose-1, 6-biphosphate aldolase to class I and class II aldolases," *Mol Microbiol.*
X53993	dapA	L-2, 3-dihydrodipicolinate synthetase (EC 4 2.1.52)	Bonnassie, S. et al. "Nucleic sequence of the dapA gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 18(21):6421 (1990)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS Microbiol. Lett* , 66:299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the Corynebacterium glutamicum lysA gene," *Mol Microbiol* , 4(11):1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component 1	Heery, D M. et al. "Nucleotide sequence of the Corynebacterium glutamicum trpE gene," *Nucleic Acids Res.*, 18(23):7138 (1990)
X56037	thrC	Threonine synthase	Han, K.S. et al. "The molecular structure of the Corynebacterium glutamicum threonine synthase gene," *Mol Microbiol* , 4(10).1693-1702 (1990)
X56075	attB-related site	Attachment site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS Microbiol. Lett.*, 66:299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit, Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J et al. "Genetic and biochemical analysis of the Aspartokinase from Corynebacterium glutamicum," *Mol Microbiol.*, 5(5).1197-1204 (1991); Kalinowski, J. et al "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspertate beta-semialdehyde dehydrogenase gene asd in Corynebacterium glutamicum," *Mol Gen Genet* , 224(3):317-324 (1990)
X59403	gap,pgk; tpi	Glyceraldehyde-3-phosphate: phosphoglycerate kinase, triosephosphate isomerase	Eikmanns, B.J "Identification, sequence analysis, and expression of a Corynebacterium glutamicum gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomeras," *J. Bacteriol* , 174(19):6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormunn, E.R et al. "Molecular analysis of the Corynebacterium glutamicum gdh gene encoding glutamate dehydrogenase," *Mol Microbiol.*, 6(3):317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A.H. et al. "Molecular analysis of the Corynebacterium glutamicum lysI gene involved in lysine uptake," *Mol Microbiol* , 5(12).2995-3005 (1991)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X66078	cop1	Ps1 protein	Joliff, G. et al "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of Corynebacterium glutamicum: The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," *Mol. Microbiol*, 6(16).2349-2362 (1992)
X66112	gli	Citrate synthase	Eikmanns, B.J. et al. "Cloning sequence, expression and transcriptional analysis of the Corynebacterium glutamicum gltA gene encoding citrate synthase," *Microbiol*, 140:1817-1828 (1994)
X67737	dapB	Dihydrodipicolinate reductase	
X69103	csp2	Surface layer protein PS2	Peyret, J.L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in Corynebacterium glutamicum," *Mol Microbiol*, 9(1):97-109 (1993)
X69104		IS3 related insertion element	Bonamy, C et al. "Identification of IS1206, a Corynebacterium glutamicum IS3-related insertion sequence and phylogenetic analysis," *Mol. Microbiol*, 14(3):571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in Corynebacterium glutamicum: enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," *Appl Environ. Microbiol*, 60(1).133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B.J. et al. "Cloning sequence analysis, expression, and inactivation of the Corynebacterium glutamicum icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," *J. Bacteriol*, 177(3).774-782 (1995)
X72855	GDHA	Glutamate dehydrogenase (NADP+)	
X75083, X70584	mtrA	5-methyltryptophan resistance	Heery, D.M. et al. "A sequence from a tryptophan-hyperproducing strain of Corynebacterium glutamicum encoding resistance to 5-methyltryptophan," *Biochem Biophys Res. Commun*, 201(3):1255-1262 (1994)
X75085	recA		Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of Corynebacterium glutamicum and Brevibacterium lactofermentum," *Appl Microbiol Biotechnol*, 42(4).575-580 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Reinscheid, D.J. et al. "Characterization of the isocitrate lyase gene from Corynebacterium glutamicum and biochemical analysis of the enzyme," *J. Bacteriol*, 176(12):3474-3483 (1994)
X76875		ATPase beta-subunit	Ludwig, W et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from Corynebacterium glutamicum," *DNA Seq.*, 4(6):403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D.J. et al. "Malate synthase from Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase: sequence analysis," *Microbiology*, 140:3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F.A. et al. "Phylogenetic analysis of the genera Rhodococcus and Norcardia and evidence for the evolutionary origin of the genus Norcardia from within the radiation of Rhodococcus species," *Microbiol.*, 141:523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronemeyer, W. et al "Structure of the gluABCD cluster encoding the glutamate uptake system of Corynebacterium glutamicum," *J Bacteriol*, 177(5):1152-1158 (1995)
X81379	dapE	Succinyldiaminopimelate desuccinylase	Wehrmann, A et al "Analysis of different DNA fragments of Corynebacterium glutamicum complementing dapE of Escherichia coli," *Microbiology*, 40:3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R et al. "Phylogeny of the genus Corynebacterium deduced from analyses of small-subunit ribosomal DNA sequences," *Int. J Syst Bacteriol*, 45(4):740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol*, 177(24):7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J Bacteriol*, 177(24):7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences," *Int. J. Syst Bacteriol.*, 45(4):724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehrmann, A. et al. "Functional analysis of sequences adjacent to dapE of Corynebacterium glutamicumproline reveals the presence of aroP, which encodes the aromatic amino acid transporter," *J Bacteriol*, 177(20):5991-5993 (1995)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X86157	argB, argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway," *Microbiology*, 142:99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase, acetate kinase	Reinscheid, D.J. et al. "Cloning, sequence analysis, expression and inactivation of the Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase and acetate kinase," *Microbiology*, 145:503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," *J. Bacteriol.*, 178(7):1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90363		Promoter fragment F45	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R.M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of Corynebacterium glutamicum," *J. Biol. Chem.*, 271(10):5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H. et al. "Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine," *J. Bacteriol.*, 178(17):5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of Corynebacterium glutamicum, encoding two enzymes involved in L-lysine synthesis," *Biotechnol. Lett.*, 19:1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein, Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum," *Mol. Microbiol.*, 22(5):815-826 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X96580	panB, panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-alanine ligase; xylulokinase	Sahm, H. et al. "D-pantothenate synthesis in Corynebacterium glutamicum and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," *Appl. Environ. Microbiol.*, 65(5):1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	
X99289		Elongation factor P	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer Brevibacterium lactofermentum (Corynebacterium glutamicum ATCC 13869)," *Gene*, 198:217-222 (1997)
Y00140	thrB	Homoserine kinase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(9):3922 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 15(9):3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(24):10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O.P. et al. "Nucleotide sequence and fine structural analysis of the Corynebacterium glutamicum hom-thrB operon," *Mol. Microbiol.*, 2(1):63-72 (1988)
Y08964	murC, ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase, division initiation protein or cell division protein; cell division protein	Honrubia, M.P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from Brevibacterium lactofermentum," *Mol. Gen. Genet.*, 259(1):97-104 (1998)
Y09163	putP	High affinity proline transport system	Peter, H. et al. "Isolation of the putP gene of Corynebacterium glutamicumproline and characterization of a low-affinity uptake system for compatible solutes," *Arch. Microbiol.*, 168(2):143-151 (1997)
Y09548	pyc	Pyruvate carboxylase	Peters-Wendisch, P.G. et al. "Pyruvate carboxylase from Corynebacterium glutamicum: characterization, expression and inactivation of the pyc gene," *Microbiology*, 144:915-927 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Patek, M. et al. "Analysis of the leuB gene from Corynebacterium glutamicum," *Appl. Microbiol. Biotechnol.*, 50(1):42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of corynephage Phi-16: The construction of an integration vector," *Microbiol.*, 145:539-548 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of Corynebacterium glutamicum glnA gene encoding glutamine synthetase I," *FEMS Microbiol Lett.*, 154(1):81-88 (1997)
Y16642	lpd	Dihydrolipoamide dehydrogenase	
Y18059		Attachment site Corynephage 304L	Moreau, S. et al "Analysis of the integration functions of φ304L. An integrase module among corynephages," *Virology*, 255(1):150-159 (1999)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminopimelate decarboxylase (partial)	Oguiza, J.A. et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum. Regulation of argS-lysA cluster expression by arginine," *J Bacteriol.*, 175(22):7356-7362 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of Brevibacterium lactofermentum encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," *J. Bacteriol.*, 175(9):2743-2749 (1993)
Z29563	thrC	Threonine synthase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," *Appl Environ Microbiol.*, 60(7)2209-2219 (1994)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	
Z49822	sigA	SigA sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum. Characterization of sigA and sigB," *J. Bacteriol.*, 178(2):550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Oguiza, J.A. et al "The galE gene encoding the UDP-galactose 4-epimerase of Brevibacterium lactofermentum is coupled transcriptionally to the dmdR gene," *Gene*, 177.103-107 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	Oguiza, J.A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol.*, 178(2):550-553 (1996)
Z66534		Transposase	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of Brevibacterium lactofermentum ATCC 13869," *Gene*, 170(1) 91-94 (1996)

[1] A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

TABLE 3: Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	Species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
Brevibacterium	ammoniagenes	21054							
Brevibacterium	ammoniagenes	19350							
Brevibacterium	ammoniagenes	19351							
Brevibacterium	ammoniagenes	19352							
Brevibacterium	ammoniagenes	19353							
Brevibacterium	ammoniagenes	19354							
Brevibacterium	ammoniagenes	19355							
Brevibacterium	ammoniagenes	19356							
Brevibacterium	ammoniagenes	21055							
Brevibacterium	ammoniagenes	21077							
Brevibacterium	ammoniagenes	21553							
Brevibacterium	ammoniagenes	21580							
Brevibacterium	ammoniagenes	39101							
Brevibacterium	butanicum	21196							
Brevibacterium	divaricatum	21792	P928						
Brevibacterium	flavum	21474							
Brevibacterium	flavum	21129							
Brevibacterium	flavum	21518							
Brevibacterium	flavum			B11474					
Brevibacterium	flavum			B11472					
Brevibacterium	flavum	21127							
Brevibacterium	flavum	21128							
Brevibacterium	flavum	21427							
Brevibacterium	flavum	21475							
Brevibacterium	flavum	21517							
Brevibacterium	flavum	21528							
Brevibacterium	flavum	21529							
Brevibacterium	flavum			B11477					

Brevibacterium	flavum			B11478					
Brevibacterium	flavum	21127							
Brevibacterium	flavum			B11474					
Brevibacterium	healii	15527							
Brevibacterium	ketoglutamicum	21004							
Brevibacterium	ketoglutamicum	21089							
Brevibacterium	ketosoreductum	21914							
Brevibacterium	lactofermentum				70				
Brevibacterium	lactofermentum				74				
Brevibacterium	lactofermentum				77				
Brevibacterium	lactofermentum	21798							
Brevibacterium	lactofermentum	21799							
Brevibacterium	lactofermentum	21800							
Brevibacterium	lactofermentum	21801							
Brevibacterium	lactofermentum			B11470					
Brevibacterium	lactofermentum			B11471					
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	21420							
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	31269							
Brevibacterium	linens	9174							
Brevibacterium	linens	19391							
Brevibacterium	linens	8377							
Brevibacterium	paraffinolyticum					11160			
Brevibacterium	spec.						717.73		
Brevibacterium	spec.						717.73		
Brevibacterium	spec.	14604							
Brevibacterium	spec.	21860							
Brevibacterium	spec.	21864							
Brevibacterium	spec.	21865							
Brevibacterium	spec.	21866							
Brevibacterium	spec.	19240							

Corynebacterium	acetoacidophilum	21476							
Corynebacterium	acetoacidophilum	13870							
Corynebacterium	acetoglutamicum			B11473					
Corynebacterium	acetoglutamicum			B11475					
Corynebacterium	acetoglutamicum	15806							
Corynebacterium	acetoglutamicum	21491							
Corynebacterium	acetoglutamicum	31270							
Corynebacterium	acetophilum			B3671					
Corynebacterium	ammoniagenes	6872						2399	
Corynebacterium	ammoniagenes	15511							
Corynebacterium	fujiokense	21496							
Corynebacterium	glutamicum	14067							
Corynebacterium	glutamicum	39137							
Corynebacterium	glutamicum	21254							
Corynebacterium	glutamicum	21255							
Corynebacterium	glutamicum	31830							
Corynebacterium	glutamicum	13032							
Corynebacterium	glutamicum	14305							
Corynebacterium	glutamicum	15455							
Corynebacterium	glutamicum	13058							
Corynebacterium	glutamicum	13059							
Corynebacterium	glutamicum	13060							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum	21513							
Corynebacterium	glutamicum	21526							
Corynebacterium	glutamicum	21543							
Corynebacterium	glutamicum	13287							
Corynebacterium	glutamicum	21851							
Corynebacterium	glutamicum	21253							
Corynebacterium	glutamicum	21514							
Corynebacterium	glutamicum	21516							
Corynebacterium	glutamicum	21299							

Corynebacterium	glutamicum	21300							
Corynebacterium	glutamicum	39684							
Corynebacterium	glutamicum	21488							
Corynebacterium	glutamicum	21649							
Corynebacterium	glutamicum	21650							
Corynebacterium	glutamicum	19223							
Corynebacterium	glutamicum	13869							
Corynebacterium	glutamicum	21157							
Corynebacterium	glutamicum	21158							
Corynebacterium	glutamicum	21159							
Corynebacterium	glutamicum	21355							
Corynebacterium	glutamicum	31808							
Corynebacterium	glutamicum	21674							
Corynebacterium	glutamicum	21562							
Corynebacterium	glutamicum	21563							
Corynebacterium	glutamicum	21564							
Corynebacterium	glutamicum	21565							
Corynebacterium	glutamicum	21566							
Corynebacterium	glutamicum	21567							
Corynebacterium	glutamicum	21568							
Corynebacterium	glutamicum	21569							
Corynebacterium	glutamicum	21570							
Corynebacterium	glutamicum	21571							
Corynebacterium	glutamicum	21572							
Corynebacterium	glutamicum	21573							
Corynebacterium	glutamicum	21579							
Corynebacterium	glutamicum	19049							
Corynebacterium	glutamicum	19050							
Corynebacterium	glutamicum	19051							
Corynebacterium	glutamicum	19052							
Corynebacterium	glutamicum	19053							
Corynebacterium	glutamicum	19054							

Corynebacterium	glutamicum	19055							
Corynebacterium	glutamicum	19056							
Corynebacterium	glutamicum	19057							
Corynebacterium	glutamicum	19058							
Corynebacterium	glutamicum	19059							
Corynebacterium	glutamicum	19060							
Corynebacterium	glutamicum	19185							
Corynebacterium	glutamicum	13286							
Corynebacterium	glutamicum	21515							
Corynebacterium	glutamicum	21527							
Corynebacterium	glutamicum	21544							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum			B8183					
Corynebacterium	glutamicum			B8182					
Corynebacterium	glutamicum			B12416					
Corynebacterium	glutamicum			B12417					
Corynebacterium	glutamicum			B12418					
Corynebacterium	glutamicum			B11476					
Corynebacterium	glutamicum	21608							
Corynebacterium	lilium		P973						
Corynebacterium	nitrilophilus	21419				11594			
Corynebacterium	spec.		P4445						
Corynebacterium	spec.		P4446						
Corynebacterium	spec.	31088							
Corynebacterium	spec.	31089							
Corynebacterium	spec.	31090							
Corynebacterium	spec.	31090							
Corynebacterium	spec	31090							20145
Corynebacterium	spec.	15954							
Corynebacterium	spec.	21857							
Corynebacterium	spec.	21862							
Corynebacterium	spec.	21863							

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Coleccion Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4th edn), World federation for culture collections world data center on microorganisms, Saimata, Japen.

>>RXA02436-amino acid sequence
(1-561, translated) 187 residues
VVGISLDVVM MGVMTSKTAT AILHTNRGDI TIDLFGNHAP ETVANFVGLA QGTKDYQSAN AQGDSEGPFY
NGSVFHRVID GFMIQGEDPT GTGRGGPGYT FADEFHPELR FDRAYLLAMA NAGPGTNGSQ FFITVTPTPH
LNNAHTIFGE VTDAESQKVV DAIATTATDR YDRPADAVVI ESVEITA
>RXA02436-nucleotide sequence A: upstream
ATCCACTACTAACTTGAGGTTGCCAATCGTGCTCATATTCAACTCTCCTTAAATTGGTCGCGTTCTTCTAAGACCCC
ACATTAGTAAAAGCATGGGACAT
>RXA02436-nucleotide sequence B: coding region
GTGGTTGGCATTTCGCTCGATGTAGTTATGATGGGCGTTATGACTTCTAAGACCGCAACTGCGATTCTGCATACAAA
TCGCGGAGACATCACCATCGACCTGTTCGGCAACCACGCTCCAGAGACCGTCGCTAACTTCGTTGGCCTGGCACAGG
GCACCAAGGACTACCAGTCCGCAAACGCTCAGGGCGACAGCGAAGGTCCGTTCTACAACGGATCTGTCTTCCACCGC
GTCATCGACGGTTTCATGATCCAGGGTGAAGACCCAACCGGCACCGGCCGTGGCGGCCCTGGCTACACCTTCGCTGA
TGAATTCCACCCAGAGCTGCGCTTCGACCGCGCATACCTGCTGGCAATGGCAAATGCGGGCCCAGGCACCAACGGTT
CCCAGTTCTTCATCACTGTGACCCCAACCCCTCACCTGAACAACGCTCACACCATCTTCGGTGAGGTCACTGACGCT
GAGTCTCAGAAGGTTGTGGATGCAATTGCAACCACCGCAACCGATCGTTACGACCGCCCAGCTGACGCAGTTGTCAT
CGAGTCTGTAGAGATCACCGCG
>RXA02436-nucleotide sequence C: downstream
TAACAGCCACCTCTACGTACACT

>>RXA02323-amino acid sequence
(1-924, translated) 308 residues
MSSHDLVDVV VVGAGAAGLA AAVALGRSLR SVIVIDAGQP RNSYAHAAHN VLGQEGIAPA ELLEKGRAEA
RSYGVTIAPG RVAKVERTGS TFAITLDDAS LLHSRRIILA HGAVDDLPEV EGLSDFWGTK VLHCAYCHGF
EARDSEIVVV GTSPMAAHQA LMFSQLSKTV SLVGTIDIDE QTSESLDSAG VKVLGTNAVR VSAEGDGLSV
ELSEGDHLSC DNIVVASRPL VDGTLYTQLG GQMEENPMGR FIPGTQTGRT PIEGVWAAGN AQAPMAMVYG
SAAQGVMAGA EINFDLILED ISVASAQS
>RXA02323-nucleotide sequence A: upstream
GGAATCATTTTCAATAGAGTCGACGCAAGTGTACACTTCTTAATGGAAATTGTTTTCAATAAAGTCAAGTTTTTTGA
CCTTCGCTTTTTAGGAGCACCCC
>RXA02323-nucleotide sequence B: coding region
ATGTCATCTCATGATCTCGTTGACGTAGTTGTCGTCGGCGCCGGCGCTGCAGGTCTCGCCGCCGCTGTCGCGCTCGG
CCGCTCACTGCGCAGTGTCATCGTCATCGACGCTGGTCAACCCCGTAACAGCTATGCGCACGCTGCTCACAATGTCC
TCGGCCAGGAAGGCATTGCGCCCGCCGAGCTGCTGGAAAAAGGCCGCGCCGAAGCGCGTTCCTATGGCGTCACCATT
GCGCCCGGGCGCGTAGCAAAAGTTGAGCGCACCGGTTCCACCTTCGCCATAACGCTTGACGACGCCTCCCTCCTTCA
CTCTCGGCGCATCATTTTGGCCCACGGCGCCGTTGACGATCTGCCAGAGGTAGAAGGACTGTCAGATTTTTGGGGAA
CCAAAGTGTTGCACTGCGCTTACTGCCACGGCTTTGAGGCCCGCGATTCTGAAATCGTCGTGGTGGGTACCTCGCCC
ATGGCTGCGCACCAAGCGTTGATGTTCTCGCAGTTGTCCAAAACTGTCAGCTTGGTGGGCACGATCGACATTGATGA
ACAAACCAGCGAGAGCCTAGATAGTGCTGGAGTAAAAGTGTTGGGCACCAATGCGGTGCGCGTATCCGCCGAAGGTG
ATGGCCTGTCTGTGGAACTGTCCGAAGGCGATCATTTAAGCTGCGACAACATCGTGGTGGCATCTCGTCCACTGGTG
GATGGCACGCTGTACACCCAACTTGGTGGTCAGATGGAAGAAAACCCGATGGGCAGGTTCATTCCAGGTACCCAAAC
CGGGCGCACTCCTATTGAAGGTGTGTGGGCTGCCGGAAACGCGCAAGCTCCCATGGCGATGGTCTATGGTTCCGCTG
CTCAAGGCGTGATGGCTGGAGCAGAGATCAACTTTGATCTGATCCTGGAAGATATTTCCGTGGCAAGCGCGCAGAGC
>RXA02323-nucleotide sequence C: downstream
TAAACTGCGTGAGGTTGTGGCCT

>>RXA02047-amino acid sequence
(1-192, translated) 64 residues
WDDRVHPAHA RLFAQALLDA GQAVDYYENT EGGHAGAADN KQTAFVESLI YTWIEKTLDQ QGSI
>RXA02047-nucleotide sequence B: coding region
TGGGACGACCGCGTCCACCCCGCGCACGCGCGCCTTTTTGCTCAAGCTTTGCTTGATGCGGGCCAGGCCGTGGATTA
CTACGAAAACACCGAGGGCGGCCATGCCGGCGCGGCGGATAACAAGCAGACCGCGTTTGTGGAATCGCTGATCTACA
CCTGGATCGAAAAGACTTTGGATCAGCAGGGTAGCATT
>RXA02047-nucleotide sequence C: downstream
TAATACCTATGATTATGCGAAGG

>>RXA01863-amino acid sequence
(1-807, translated) 269 residues
MNSNLTLP QDRYLDIAEY IDVMHISHNW GTTDEFANVG FGAMKKQPPL KAKLKLYEQM ISNARTLSEQ
GMFVSAETML NQSTLPHLRK IHQEVVHDMK CSRHEIHPMY PADFASQLNV LTLAEMKKTI HDILDFRDED
IWMLFGTLPV FPCLKDDEDQ KLLSRLRNAN NVTTRNDPDG RSRLNVNVFT GNVIVTDFGD ETGTISNIQK
DKLTDVFDKW LSSDLAKSLN CHCSEFSCLG PNVLVKNMYY PNMDFKDNER HMHKQPQIIQ F
>RXA01863-nucleotide sequence A: upstream
TTAGAAATGTTGTTAACCTCTATTAAAGTATGCACATCATCGAGGTATATATACACAA
>RXA01863-nucleotide sequence B: coding region
ATGAATTCAAACCTAACATTGCCTCAAGATCGTTATTTAGATATTGCTGAATATATCGATGTTATGCATATCTCACA
TAACTGGGGAACAACTGATGAATTCGCAAATGTTGGCTTTGGCGCAATGAAGAAGCAACCACCGTTAAAAGCTAAGT
TAAAATTATATGAACAAATGATTTCGAATGCACGTACATTATCAGAACAAGGAATGTTTGTATCTGCGGAAACAATG
CTCAATCAAAGTACGCTACCACATTTACGAAAAATACATCAAGAAGTCGTTCATGATATGAAATGTAGCAGACACGA
GATTCACCCTATGTATCCAGCTGACTTTGCAAGTCAATTAAATGTGTTAACTCTAGCGGAAATGAAAAAGACAATTC
ATGATATATTGGATTTCAGAGATGAAGATATTTGGATGTTATTTGGTACTTTGCCTGTGTTTCCATGCTTAAAGGAT
GATGAAGATCAAAAGTTACTATCACGTTTAAGAAATGCTAACAATGTAACGACTAGAAATGACCCGGATGGCCGTAG
TCGTTTAAATGTCAATGTATTTACAGGTAATGTAATCGTAACTGATTTCGGAGATGAAACAGGTACAATTTCGAATA
TACAAAAAGATAAATTAACAGATGTATTTGATAAATGGTTATCCTCTGATCTTGCTAAATCATTAAATTGTCATTGT
TCCGAGTTTAGTTGTTTAGGACCAAATGTTCTTGTTAAAAATATGTACTATCCGAATATGGATTTTAAAGATAATGA
GCGTCATATGCACAAACAACCACAAATTATACAATTT
>RXA01863-nucleotide sequence C: downstream
TAAAAACTCTTAATTATGCGGAG

>>RXA01841-amino acid sequence
(1-363, translated) 121 residues
MEEGEEISLS DFEGEVVVLN AWGQWCAPCR AEVDDLQLVQ ETLDPLGGTV LGINVRDYNQ TIAQDFKLDN
AVTYPSIYDP PFRIAAALGG VPTSVIPTTI VLDRSHRPAA VFLREVTALS G
>RXA01841-nucleotide sequence A: upstream
TCGGGGGGCACCTTCCAGTTCCACTCACCAGAAGGCCAACCGAGATTATCTACGCAAAGGAGGAACGTGCCCCGCTG
CGGACTTCTCCGGCCCGTCGCTG
>RXA01841-nucleotide sequence B: coding region
ATGGAGGAGGGTGAGGAGATCAGCCTGTCTGATTTTGAAGGCGAGGTCGTCGTCCTCAACGCTTGGGGCCAGTGGTG
TGCACCGTGTCGGGCGGAAGTCGATGACCTGCAGCTTGTCCAGGAGACTCTCGACCCCCTCGGTGGCACGGTGCTGG
GCATCAACGTCCGTGACTACAACCAGACCATCGCCCAGGACTTCAAACTCGACAACGCGGTGACCTATCCCTCGATC
TACGACCCGCCGTTTCGTATCGCTGCGGCCCTGGGTGGGGTGCCGACCTCGGTCATCCCGACCACCATTGTCCTGGA
CCGAAGCCACCGCCCGGCCGCGGTGTTCCTGAGGGAGGTCACCGCGTTGTCTGGT
>RXA01841-nucleotide sequence C: downstream
TAGTCCTGGGAGGTGATGAGCTC

>>RXA01837-amino acid sequence
(1-393, translated) 131 residues
SEGYYNDTVC HRITTSGIYV LQCGDPSSTG AGGPGFSFAN EYPTDEATDL TTPVIYERGT IAMANAGADT
NGSQFFLNYE DSPLAPNYTY FGQITEEGLA TLDAIAEVGT EGGTGDGAPA QEVRIESAAV A
>RXA01837-nucleotide sequence B: coding region
TCGGAGGGCTACTACAACGATACTGTCTGCCACCGCATCACCACCTCTGGCATTTACGTTCTCCAGTGCGGCGATCC
AAGCAGCACCGGCGCAGGCGGCCCAGGGTTCAGCTTCGCCAACGAATACCCAACCGACGAAGCAACTGACCTAACCA
CCCCAGTCATCTACGAGCGCGGCACCATCGCCATGGCCAACGCTGGCGCTGACACCAACGGCTCCCAGTTCTTCCTC
AACTACGAGGATTCCCCACTGGCACCGAACTACACCTACTTCGGCCAGATCACCGAAGAAGGCCTTGCAACCCTCGA
CGCCATCGCAGAAGTTGGCACTGAAGGTGGAACCGGCGACGGAGCACCAGCGCAAGAGGTTCGCATTGAATCCGCAG
CTGTTGCG
>RXA01837-nucleotide sequence C: downstream
TAAGTTCTAAGCCCCTCTTCTTT

>>RXA02174-amino acid sequence
(1-354, translated) 118 residues
MEKPQIELPV GPAPEDLVIS DIIVGEGAEA RPGGEVEVHY VGVDFETGEE FDSSWDRGQT SQFPLNGLIA
GWQEGIPGMK VGGRRQLTIP PEAAYGPEGS GHPLSGRTLV FIIDLISA
>RXA02174-nucleotide sequence A: upstream
CTTCGGCCGAAGCCACTTCGTCTGTCATAATGACAGGGATGGTTTCGGCCGTTTTTGCATGAAACCAAAAAATACGA
TTTTCAAGGAGCATGTACAGCAC
>RXA02174-nucleotide sequence B: coding region
ATGGAAAAGCCACAGATTGAGCTACCGGTCGGTCCAGCACCGGAAGATCTCGTAATCTCTGACATCATCGTTGGCGA
AGGAGCAGAAGCCCGCCCAGGTGGAGAAGTTGAGGTCCACTATGTGGGCGTTGACTTTGAAACCGGCGAGGAGTTTG
ACTCTTCCTGGGATCGTGGACAGACCAGCCAGTTCCCACTCAACGGCCTCATTGCAGGTTGGCAAGAGGGAATTCCA
GGCATGAAGGTCGGCGGACGTCGTCAGCTGACCATTCCGCCAGAGGCTGCTTACGGCCCTGAGGGTTCCGGCCACCC
ACTGTCTGGCCGTACCCTGGTGTTCATCATCGATTTGATCAGCGCA
>RXA02174-nucleotide sequence C: downstream
TAATTTTCTTTACTGCGCTAAAC

>>RXA01072-amino acid sequence
(1-231, translated) 77 residues
MAITVYTKPA CVQCNATKKA LDRAGLEYDL VDISLDEEAR EYVLALGYLQ APVVVADGSH WSGFRPERIR
EMATAAA
>RXA01072-nucleotide sequence A: upstream
GTGTCGCCTTGAGCGATGCAAAGACTTTAAAAACAGATCTAAAAAAGTAAGATGAGGCATTAGCGTCAACCGCTAAA
GCCGATAAGGGAAGGTCCAAAAA
>RXA01072-nucleotide sequence B: coding region
ATGGCAATCACCGTTTACACCAAGCCAGCTTGCGTCCAGTGCAATGCCACCAAGAAGGCCCTCGACCGCGCTGGTCT
TGAGTATGACCTCGTTGATATCAGCCTTGATGAAGAGGCACGTGAGTACGTCCTCGCACTTGGCTACCTGCAGGCAC
CAGTTGTCGTTGCAGATGGCTCCCACTGGTCCGGTTTCCGCCCAGAGCGCATCCGTGAAATGGCAACCGCAGCTGCC
>RXA01072-nucleotide sequence C: downstream
TAAACTGCACTTCGTGGCACGAC

>>RXA00824-amino acid sequence
(1-558, translated) 186 residues
MTSSAKWSIV GVVAILAVIV ALIPQLVGGE SAEEAQGETS TSKITTRPDC VASGAAGVDL PCLGGANGVG
NELATVVNLW AWWCEPCRAE LPIFDEFATT HPELNVIGVH ADQNAANGAA LLEDLGVNLA SYQDDSNLFA
GTLGLPGVVP ITIVVSPDGN VVDTFPQPFE TIDDLETAVA GALQNA
>RXA00824-nucleotide sequence A: upstream
TTTGGTTTGGAGGGGCCGTCAGATCCATTTGAGGCGCAAAAACTCATTAAAAGTGATGATAGGGAGCACCTGCTGAA
AATGGCAGGAATGTAGAAAACAA
>RXA00824-nucleotide sequence B: coding region
ATGACAAGCAGTGCAAAGTGGTCCATCGTTGGAGTTGTCGCCATCCTGGCTGTGATCGTTGCGTTAATCCCGCAGCT
TGTGGGTGGAGAAAGCGCAGAGGAAGCGCAGGGGGAGACGTCGACAAGCAAAATTACGACGCGTCCTGACTGCGTGG
CCTCTGGCGCGGCGGGTGTGGATCTGCCCTGCTTGGGCGGCGCCAACGGCGTCGGCAACGAGCTGGCCACCGTGGTG
AATCTGTGGGCGTGGTGGTGCGAACCGTGCCGCGCGGAGCTGCCGATTTTTGATGAATTCGCCACCACCCACCCCGA
ACTCAACGTCATTGGCGTGCATGCAGACCAAAACGCAGCCAACGGCGCCGCACTCCTTGAGGATCTGGGCGTGAATC
TTGCAAGCTACCAAGACGATTCCAACCTGTTCGCAGGCACCCTTGGGCTGCCGGGCGTCGTGCCGATCACCATCGTG
GTTTCTCCAGACGGCAATGTAGTGGACACCTTCCCGCAGCCTTTCGAAACCATCGATGACCTCGAAACCGCTGTGGC
AGGGGCGCTGCAGAATGCG
>RXA00824-nucleotide sequence C: downstream
TAACTACCCTGATTTACCGCATG

>>RXA00568-amino acid sequence
(1-1347, translated) 449 residues
VKSSVEKLSD TRSKITVEVP FSELKPEIDQ AYAALAQQVQ IPGFRKGKAP RQLIDARFGR GAVLEQVVND
MLPNRYAQAI EAEGIKAIGQ PNVEVTKIED NELVEFVAEV DVRPEFELPK FEDITVEVPA IKADEEAIEA
ELETLRARFS TLKDHNHKLK KGEFVTINIT ASIDGEKIEE ATTEGLSYEI GSDDLIDGLD KALIGAKKDE
TVEFTSELAN GEHKGKEAQI SVEITATKQR ELPELDDEFA QLASEFDTIE ELRESTVSDV EAKQKNEQAA
AIRDEVLAAA LGEADFALPQ SIVDEQAHSQ LHQLLGELAH DDAALNSLLE AQGTTREEFD KKNVEDAEKA
VRTQLFLDTL SEVEEPEVSQ QELTDHILFT AQSYGMDPNQ FIGQLQQSGQ IANLFSDVRR GKALAQAICR
VNVKDSEGNE IDPKEYFGEE EVAETESEA
>RXA00568-nucleotide sequence A: upstream
GTTTCGGACGACGCGAGAAATCGCATTAATCGTCGGAAACCGGGCTTGTTTTTGTAATATCTGAAACTTTCCCTTTC
CCGATCATCCAGGAGATTTACTC
>RXA00568-nucleotide sequence B: coding region
GTGAAGAGTTCTGTCGAGAAGCTGAGCGACACCCGTTCAAAGATCACCGTTGAGGTTCCATTTTCTGAACTGAAGCC
AGAGATCGACCAGGCATACGCCGCTCTAGCGCAGCAAGTCCAGATCCCTGGTTTCCGTAAGGGCAAGGCACCGCGTC
AGCTTATCGACGCACGCTTCGGCCGTGGTGCGGTTCTGGAGCAGGTTGTCAACGACATGCTTCCTAACCGCTACGCA
CAGGCAATCGAAGCTGAGGGCATCAAGGCAATCGGCCAGCCTAACGTAGAGGTCACCAAGATCGAAGACAACGAGCT
CGTTGAGTTCGTCGCTGAGGTTGACGTTCGCCCAGAGTTCGAGCTTCCTAAGTTCGAGGACATCACTGTTGAGGTCC
CAGCTATCAAGGCTGACGAAGAGGCAATCGAAGCAGAGCTCGAGACCCTGCGTGCACGTTTCTCCACCTTGAAGGAT
CACAACCACAAGCTGAAGAAGGGTGAGTTCGTCACCATCAACATCACCGCAAGCATTGACGGTGAGAAGATTGAAGA
GGCAACCACTGAGGGTCTGTCCTACGAAATCGGATCTGACGATCTGATTGACGGCCTGGACAAGGCTCTGATCGGCG
CTAAGAAGGATGAAACCGTAGAGTTCACCTCTGAGCTGGCAAACGGCGAGCACAAGGGCAAGGAAGCTCAAATCAGC
GTTGAGATCACCGCAACCAAGCAGCGCGAGCTGCCTGAGCTGGATGATGAGTTCGCACAGCTGGCTTCTGAGTTCGA
CACCATCGAAGAGCTTCGTGAGTCCACCGTGTCTGACGTTGAGGCTAAGCAGAAGAACGAGCAGGCTGCTGCAATCC
GCGACGAAGTTCTCGCTGCGGCTCTTGGCGAGGCTGACTTCGCTCTGCCACAGTCCATCGTTGACGAGCAGGCACAC
TCCCAGCTGCACCAGCTCCTCGGCGAGCTTGCACACGACGATGCTGCACTGAACTCCCTCCTTGAGGCTCAGGGCAC
CACTCGTGAAGAGTTCGACAAGAAGAACGTCGAAGATGCTGAGAAGGCTGTTCGCACCCAGCTGTTCCTGGACACCC
TCTCTGAGGTTGAGGAGCCTGAGGTTTCCCAGCAGGAGCTCACCGACCACATCCTGTTCACCGCACAGTCTTACGGC
ATGGACCCAAACCAGTTCATCGGTCAGCTGCAGCAGTCCGGCCAGATCGCGAACCTCTTCTCCGACGTTCGCCGTGG
CAAGGCTCTTGCACAGGCTATCTGCCGCGTAAACGTGAAGGACTCCGAGGGTAACGAGATCGACCCTAAGGAATACT
TCGGTGAAGAAGAAGTAGCTGAGACTGAGTCTGAAGCT
>RXA00568-nucleotide sequence C: downstream
TAAAAACTTTAAAGAAATAACGC

Claims

1. An isolated nucleic acid molecule from *Corynebacterium glutamicum* encoding an SES protein, or a portion thereof.

2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule encodes an SES protein involved in the production of a fine chemical.

3. An isolated *Corynebacterium glutamicum* nucleic acid molecule selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

4. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth in Appendix B.

5. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth in Appendix B.

6. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

7. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A.

8. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1-7 under stringent conditions.

9. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-8 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.

10. A vector comprising the nucleic acid molecule of any one of claims 1-9.

11. The vector of claim 10, which is an expression vector.

12. A host cell transfected with the expression vector of claim 11.

13. The host cell of claim 12, wherein said cell is a microorganism.

14. The host cell of claim 13, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

15. The host cell of claim 12, wherein the expression of said nucleic acid molecule results in the modulation in production of a fine chemical from said cell.

16. The host cell of claim 15, wherein said fine chemical is selected from the group consisting of: organic acids. proteinogenic and nonproteinogenic amino acids. purine and pyrimidine bases, nucleosides. nucleotides. lipids, saturated and unsaturated fatty acids, diols, carbohydrates. aromatic compounds, vitamins, cofactors, and enzymes.

17. A method of producing a polypeptide comprising culturing the host cell of claim 12 in an appropriate culture medium to. thereby, produce the polypeptide.

18. An isolated SES polypeptide from *Corynebacterium glutamicum*, or a portion thereof.

19. The polypeptide of claim 18, wherein said polypeptide is involved in the production of a fine chemical production.

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, or a portion thereof.

22. The isolated polypeptide of any of claims 18-21, further comprising heterologous amino acid sequences.

23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth in Appendix A.

24. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

25. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 12 such that the fine chemical is produced.

26. The method of claim 25. wherein said method further comprises the step of recovering the fine chemical from said culture.

27. The method of claim 25, wherein said method further comprises the step of transfecting said cell with the vector of claim 11 to result in a cell containing said vector.

28. The method of claim 25, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

29. The method of claim 25, wherein said cell is selected from the group consisting of: *Corynebacterium glutamicum. Corynebacterium herculis. Corynebacterium. lilium. Corynebacterium acetoacidophilum. Corynebacterium acetoglutamicum.*

Corynebacterium acetophilum, Corynebacterium ammoniagenes, Corynebacterium fujiokense, Corynebacterium nitrilophilus, Brevibacterium ammoniagenes, Brevibacterium butanicum, Brevibacterium divaricatum, Brevibacterium flavum, Brevibacterium healii, Brevibacterium ketoglutamicum, Brevibacterium ketosoreductum, Brevibacterium lactofermentum, Brevibacterium linens, Brevibacterium paraffinolyticum, and those strains set forth in Table 3

30. The method of claim 25, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.

31. The method of claim 25, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.

32. The method of claim 25, wherein said fine chemical is an amino acid.

33. The method of claim 32, wherein said amino acid is drawn from the group consisting of: lysine, glutamate, glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.

34. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-9.